Semester Project
Swiss Institute of Technology
EPFL

STOCHASTIC PROCESSES AND CONSTRUCTION OF BROWNIAN MOTION

Under the supervision of Prof. Thomas Mountford

November 25, 2013

Christophe-Aschkan Mery
christophe-aschkan.mery@epfl.ch

Contents

1 Finite Dimensional Distributions of Stochastic Processes **7**
- 1.1 A Projective Family . . . 7
- 1.2 Product Spaces . . . 9
- 1.3 Kolmogorov's Extension Theorem . . . 10
- 1.4 Law of a Stochastic Process . . . 12
- 1.5 Finite-Dimensional Distributions Determine the Law of a Process . . . 12

2 Continuity of Stochastic Processes **13**
- 2.1 Similarites Between Processes . . . 13
- 2.2 Continuity of Trajectories . . . 13
- 2.3 Kolmogorov's Criterion for Continuity . . . 15

3 Introduction to Brownian Motion **21**
- 3.1 Motivation . . . 21
- 3.2 Construction of a pre Brownian Motion . . . 22
- 3.3 Continuity of Paths : Existence of Brownian Motion . . . 26
- 3.4 Wiener Space . . . 26

4 Properties of Brownian Motion **29**
- 4.1 Non Differentiability of Brownian Motion . . . 29
- 4.2 Weak Markov Property . . . 31
- 4.3 Strong Markov Property . . . 32

A Preliminaries **35**
- A.1 Mesurable Functions and Random Variables . . . 35
- A.2 Convergence . . . 38
- A.3 Independence . . . 38
- A.4 Gaussian Spaces . . . 40

A Fundamental Theorem

Before we start our step-by-step construction of Brownian motion, we need to state and prove a theorem that will be one of the building blocks of the theory. It is the $\pi - \lambda$ theorem. Throughout this work, we implicitly use one of its direct consequences, giving a simple criterion for independence of the σ-algebra generated by a stochastic process from another σ-algebra on the same space.

A collection of sets $\mathcal{P}$ is said to be a **π-system** if it is closed under intersection, that is, is $A, B \in \mathcal{P}$, then $A \cap B \in \mathcal{P}$. A collection of sets $\mathcal{L}$ is said to be a **λ-system** if it satisfies : (i) $\Omega \in \mathcal{L}$. (ii) If $A, B \in \mathcal{L}$, and $A \subset B$, then $B \backslash A \in \mathcal{L}$. (iii) If $A_n \in \mathcal{L}$ and $A_n \nearrow A$, then $A \in \mathcal{L}$. As for σ-algebras, we see that an intersection of λ-systems is again a λ-system. Now, if $\mathcal{A}$ is any collection of subsets of Ω then we can define the λ-system generated by $\mathcal{A}$ by the intersection of all λ-systems containing $\mathcal{A}$. It is denoted by $l(\mathcal{A})$. These notions being defined, we can now state a very important theorem.

Theorem 1. *If $\mathcal{P}$ is a π-system, then $l(\mathcal{P}) = \sigma(\mathcal{P})$.*

Proof. First we note that, by definition, a σ-algebra is a λ-system. So, $l(\mathcal{P})$ being the intersection of all λ-systems containing $\mathcal{P}$, we must have : $l(\mathcal{P}) \subset \sigma(\mathcal{P})$. On the other side, if $l(\mathcal{P})$ is a σ-algebra, then $l(\mathcal{P}) \supset \sigma(\mathcal{P})$ and we finished the proof. Besides, a λ-system is a σ-algebra if and only if is closed under finite intersections (we can see this using the complements of the sets and the increasing limit property of a λ-system). So the only thing we have to show is that $l(\mathcal{P})$ is closed under finite intersection.
Let's fix $A \in \mathcal{P}$ and define :

$$\mathcal{L}_1 = \{B \in l(\mathcal{P}) \mid A \cap B \in l(\mathcal{P})\}.$$

Because $\mathcal{P}$ is a π-system, we have $\mathcal{P} \subset \mathcal{L}_1$. Also, $A \cap \Omega = A \in \mathcal{P} \subset l(\mathcal{P})$ so $\Omega \in \mathcal{L}_1$. Then, if $B, B' \in \mathcal{L}_1$ and $B \subset B'$, we have $A \cap (B' \backslash B) = A \cap B' \cap B^c = (A \cap B' \cap B^c) \cup (A \cap B' \cap A^c) = (A \cap B') \cap (A^c \cup B^c) = (A \cap B') \backslash (A \cap B)$. But we have $(A \cap B') \backslash (A \cap B) \in l(\mathcal{P})$ because $(A \cap B), (A \cap B') \in l(\mathcal{P})$, $(A \cap B) \subset (A \cap B')$ and $l(\mathcal{P})$ is λ-system. This gives us $(B' \backslash B) \in \mathcal{L}_1$. Now let $B_1 \subset B_2 \subset ...$ be a sequence of $\mathcal{L}_1$, then $A \cap (\bigcup B_n) = \bigcup (A \cap B_n) \in l(\mathcal{P})$ and so $\bigcup B_n \in \mathcal{L}_1$. Thus we have proved that $\mathcal{L}_1$ is a λ-system that contains $\mathcal{P}$. It must then contain $l(\mathcal{P})$. Basically, we have : $\forall A \in \mathcal{P}, \forall B \in l(\mathcal{P})$, $A \cap B \in l(\mathcal{P})$.
However, this is not exactly what we wanted, since we need to prove that $l(\mathcal{P})$ is closed under finite intersection. The final result is not far away. To conclude we can use the same idea as before and take A in $l(\mathcal{P})$ instead of $\mathcal{P}$. We then define :

$$\mathcal{L}_2 = \{B \in l(\mathcal{P}) \mid A \cap B \in l(\mathcal{P})\}.$$

Recall that we proved : $\forall B \in \mathcal{P} \subset l(\mathcal{P}), \forall A \in l(\mathcal{P})$, $A \cap B \in l(\mathcal{P})$. Applying this result here allows us to write : $\mathcal{P} \subset \mathcal{L}_2$. We can now give *exactly* the same arguments as before to show that $\mathcal{L}_2$ is a λ-system. Again, if it is a λ-system and it contains $\mathcal{P}$, then by definition of $l(\mathcal{P})$ we must have $l(\mathcal{P}) \subset \mathcal{L}_2$. So : $\forall A \in l(\mathcal{P}), \forall B \in l(\mathcal{P})$, we have $B \in \mathcal{L}_2 \implies A \cap B \in l(\mathcal{P})$. This means that $l(\mathcal{P})$ is closed under finite intersection and it concludes the proof. □

Theorem 2 ($\pi - \lambda$)**.** *If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system that contains $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Proof. If $\mathcal{L}$ is a λ-system that contains $\mathcal{P}$, then $l(\mathcal{P}) \subset \mathcal{L}$, by definition of $l(\mathcal{P})$. But the previous theorem states that $l(\mathcal{P}) = \sigma(\mathcal{P})$. Which concludes the proof.

□

Important consequence of the $(\pi - \lambda)$ theorem

Let $(X_t, t \in T)$ be a sequence of random variables. And $\mathcal{F}$ a σ-algebra on the same probability space. To proof that $\sigma(X_t, t \in T)$ and $\mathcal{F}$ are independent, it is enough to proof that for any sequence $t_1 \leq t_2 \leq ... \leq t_k$, the vector $(X_{t_1}, X_{t_2}, ..., X_{t_k})$ is independent of $\mathcal{F}$. Indeed, the class of pre-images by finite vectors of Borelian sets is a π-system which generates $\sigma(X_t, t \in T)$.

Chapter 1

Finite Dimensional Distributions of Stochastic Processes

In this chapter, we will define the notion of a stochastic process X on a space Ω and the finite dimensional distributions related to it. We will see that to each occurence $\omega \in \Omega$, we can associate a trajectory $\omega : t \longmapsto X_t(\omega)$, which is a function from $\mathbb{R}_+$ to $\mathbb{R}$. It will thus seem natural to directly consider the space of such functions for the experiment associated to X instead of Ω itself. To do so, we will have to construct a σ-algebra $\mathcal{R}^{\mathbb{R}_+}$ and a probability measure on it. This probability will be given by the Kolmogorv's extension theorem. Finally, we will show that the subset "Observing a continuous trajectory" is not an event in this probability space. Since our ultimate goal is to model the random motion of a particle in space, we will have to adress this issue at some point.

A **stochastic process** is a collection $X = (X_t, t \in T)$ of random variables on a probability space $(\Omega, \mathcal{F}, P)$, T is thought of as representing time. Usually, it is discrete (subset of $\mathbb{Z}$) or continuous (subset of $\mathbb{R}$). For theoretical purposes however, we can consider it arbitrary.

1.1 A Projective Family

Let's fix a k-tuple of "times" in T: $t_1 < t_2 < ... < t_k$. Then, the joint distribution of $(X_{t_1}, X_{t_2}, ..., X_{t_h})$ is given by :

$$\mu_{t_1 t_2 ... t_k}(H) := P[\omega \in \Omega : (X_{t_1}(\omega), X_{t_2}(\omega), ..., X_{t_k}(\omega)) \in H], \qquad \text{for } H \in \mathcal{B}(\mathbb{R}^k) \tag{1.1}$$

The probability measures $\mu_{t_1 t_2 ... t_k}$ are called the **finite dimensional distributions** (or FDD) of the stochastic process $X = (X_t, t \in T)$.

Now, let's fix $t_1 < t_2 < ... < t_k$ and let $H = H_1 \times H_2 \times ... \times H_{k-1}$ be any rectangle where $H_i \in \mathcal{B}(\mathbb{R})$. Since two events that are the same must have the same probability, (1.1) implies that :

$$\begin{aligned}
\mu_{t_1 t_2 ... t_{k-1}}(H) &= P[\omega \in \Omega : (X_{t_1}(\omega), X_{t_2}(\omega), ..., X_{t_{k-1}}(\omega)) \in H_1 \times H_2 \times ... \times H_{k-1}] \\
&= P[\omega \in \Omega : (X_{t_1}(\omega), X_{t_2}(\omega), ..., X_{t_{k-1}}(\omega), X_{t_k}(\omega)) \in H_1 \times H_2 \times ... \times H_{k-1} \times \mathbb{R}] \\
&= \mu_{t_1 t_2 ... t_{k-1} t_k}(H_1 \times H_2 \times ... \times H_{k-1} \times \mathbb{R}).
\end{aligned}$$

Also, if π is any permutation of $(1, 2, .., k)$ we have :

$$\begin{aligned}
&\mu_{t_1 t_2 ... t_k}(H_1 \times H_2 \times ... \times H_k) \\
&\qquad = P[\omega \in \Omega : (X_{t_1}(\omega), X_{t_2}(\omega), ..., X_{t_k}(\omega)) \in H_1 \times H_2 \times ... \times H_k] \\
&\qquad = P[\omega \in \Omega : (X_{t_{\pi(1)}}(\omega), X_{t_{\pi(2)}}(\omega), ..., X_{t_{\pi(k)}}(\omega)) \in H_{\pi(1)} \times H_{\pi(2)} \times ... \times H_{\pi(k)}] \\
&\qquad = \mu_{t_{\pi(1)} t_{\pi(2)} ... t_{\pi(k)}}(H_{\pi(1)} \times H_{\pi(2)} \times ... \times H_{\pi(k)}).
\end{aligned}$$

.

To summarize, finite distributions $\mu_{t_1 t_2 ... t_k}$ coming from a stochastic process $X = (X_t, t \in T)$ via (1.1) necessarily satisfy (for all rectangles $H_1 \times H_2 \times ... \times H_k$):

1. **Projection property**

$$\mu_{t_1 t_2 ... t_{k-1} t_k}(H_1 \times H_2 \times ... \times H_{k-1} \times \mathbb{R}) = \mu_{t_1 t_2 ... t_{k-1}}(H_1 \times H_2 \times ... \times H_{k-1}) \tag{1.2}$$

2. **Permutation invariance**

$$\mu_{t_1 t_2 ... t_k}(H_1 \times H_2 \times ... \times H_k) = \mu_{t_{\pi(1)} t_{\pi(2)} ... t_{\pi(k)}}(H_{\pi(1)} \times H_{\pi(2)} \times ... \times H_{\pi(k)}) \tag{1.3}$$

for any permutation π of $(1, 2, ..., k)$.

Conversersely, *Kolmogorov's existence theorem* tells us that if we are given a system of finite dimensional distributions satisfying (1.2) and (1.3), then we can construct a stochastic process having these finite dimensional distributions. The proof of this theorem will be a construction and to simplify it, we introduce some new notions.

If $\phi : \mathbb{R}^k \to \mathbb{R}^{k-1}$ is the projection $\phi(x_1, x_2, .., x_k) = (x_1, x_2, .., x_{k-1})$, then condition (1.2) reads:

$$\mu_{t_1 t_2 ... t_{k-1}}(H_1 \times H_2 \times ... \times H_{k-1}) = \mu_{t_1 t_2 ... t_k} \phi^{-1}(H_1 \times H_2 \times ... \times H_{k-1})$$

or

$$\mu_{t_1 t_2 ... t_{k-1}} = \mu_{t_1 t_2 ... t_k} \phi^{-1}. \tag{1.4}$$

Similarly, given a permutation π of $(1, 2, ..., k)$ we can defined a function $\phi_\pi : \mathbb{R}^k \to \mathbb{R}^k$ which applies this permutation to the elements of a vector. More precisely :

$$\phi_\pi(x_1, x_2, ..., x_k) = (x_{\pi^{-1}(1)}, x_{\pi^{-1}(2)}, ..., x_{\pi^{-1}(k)})$$

and

$$\phi_\pi^{-1}(x_1, x_2, ..., x_k) = (x_{\pi(1)}, x_{\pi(2)}, ..., x_{\pi(k)}).$$

With this formalism, condition (1.3) becomes :

$$\mu_{t_{\pi(1)} t_{\pi(2)} ... t_{\pi(k)}} \phi_\pi^{-1}(H_1 \times H_2 \times ... \times H_k) = \mu_{t_1 t_2 ... t_k}(H_1 \times H_2 \times ... \times H_k)$$

or

$$\mu_{t_1 t_2 ... t_k} = \mu_{t_{\pi(1)} t_{\pi(2)} ... t_{\pi(k)}} \phi_\pi^{-1}. \tag{1.5}$$

We already see that this re-writing of (1.2) and (1.3) reveals some similarities *in form* of these two conditions. And in fact, they have a common extension.

We still consider $t_1 < t_2 < ... < t_k$ fixed as before. And let $u_1 < u_2 < ... < u_m$ be any m-tuple so that each t_i is in it (necessarily $m \geq k$). There must be a permutation of $(u_1, u_2, ..., u_m)$ so that $(t_1, t_2, ..., t_k)$ is the first segment of the vector obtained after permutation. Let π be the corresponding permutation of $(1, 2, ..., m)$ that satisfies this condition that is :

$$(u_{\pi^{-1}(1)}, ..., u_{\pi^{-1}(m)}) = (t_1, t_2, ..., t_k, u_{\pi^{-1}(k+1)}, ..., u_{\pi^{-1}(m)})$$

We can now define $\psi : \mathbb{R}^m \longrightarrow \mathbb{R}^k$. This function first applies the desired permutation (of $(u_1, u_2, ..., u_m)$) to the coordinates of a vector and then projects the vector thus obtained into $\mathbb{R}^k$. Precisely, we have that $\psi(X_{u_1}, X_{u_2}, ..., X_{u_m}) = (X_{t_1}, X_{t_2}, ..., X_{t_k})$ and :

$$\begin{aligned}
&\{\omega \in \Omega : (X_{t_1}(\omega), X_{t_2}(\omega), ..., X_{t_k}(\omega)) \in H_1 \times H_2 \times ... \times H_k\} \\
&= \{\omega \in \Omega : \psi(X_{u_1}(\omega), X_{u_2}(\omega), ..., X_{u_m}(\omega)) \in H_1 \times H_2 \times ... \times H_k\} \\
&= \{\omega \in \Omega : (X_{u_1}(\omega), X_{u_2}(\omega), ..., X_{u_m}(\omega)) \in \psi^{-1}(H_1 \times H_2 \times ... \times H_k)\}
\end{aligned}$$

and so :

$$\mu_{t_1 t_2 ... t_k} = \mu_{u_1 u_2 ... u_m} \psi^{-1}. \tag{1.6}$$

which is a condition that contains both (1.4) and (1.5) as special cases. But since ψ is a coordinate permutation followed by a sequence of projections of the form $(x_1, x_2, ..., x_l) \longmapsto (x_1, x_2, ..., x_{l-1})$, it is also a consequence of these special cases.

1.2 Product Spaces

Let T be an arbitray index set, then $\mathbb{R}^T$ is the collection of all real functions on T. For instance, we could have $T = \mathbb{R}_+$ and in this case $\mathbb{R}^{\mathbb{R}_+}$ would represent all the functions $f : \mathbb{R}_+ \longrightarrow \mathbb{R}$. In particular, if $X = (X_t, t \in \mathbb{R}_+)$ is a stochastic process and if we fix $\omega \in \Omega$, we get a point of $\mathbb{R}^{\mathbb{R}_+}$:

$$\mathbb{R}^{\mathbb{R}_+} \longrightarrow \mathbb{R}$$

$$\omega \longmapsto X_t(\omega).$$

which we shall denote by ω. Thus, ω can be seen either like a realisation $\omega \in \Omega$ or like a function $\omega \in \mathbb{R}^{\mathbb{R}_+}$. In this case, ω is a specific **trajectory** of the particle modelled by X. This ambiguity doesn't make much difference in practice.
If $T = \{1, 2, ...\}$, $\mathbb{R}^T$ can be thought of as the set of infinite sequences (x_n) taking their values in $\mathbb{R}$. Finally, if $T = \{1, 2, ..., k\}$ is finite, then $\mathbb{R}^T$ can be identified to $\mathbb{R}^k$. For any $x \in \mathbb{R}^T$, x can be considered as a point in a certain space, a sequence of numbers or a function. This is equal. The mathematical object x is the same and all that changes is our *representation* of that object. For $x \in \mathbb{R}^T$ and $t \in T$ fixed, the real number associated to x and t can be written x_t or $x(t)$ depending on the choosen representation of x. Now, for each t, we can define a mapping

$$Y_t \colon \mathbb{R}^T \to \mathbb{R}$$

$$x \longmapsto Y_t(x) = x(t) = x_t.$$

The Y_t are called **coordinate mappings**. Later on, when we define a σ-algebra and a probability measure on $\mathbb{R}^T$, the Y_t will become random variables (see (3.1)) associating to a particle following a trajectory x its position $x(t)$ at time t.

We now construct a σ-algebra on $\mathbb{R}^T$. For each $H \in \mathcal{B}(\mathbb{R})$, we have

$$Y_t^{-1}(H) = \{x \in \mathbb{R}^T : Y_t(x) \in H\} = \{x \in \mathbb{R}^T : x(t) \in H\} \tag{1.7}$$

for $t \in T$. Let's consider all the sets of this form (for $H \in \mathcal{B}(\mathbb{R})$ and $t \in T$). They generate a σ-algebra on $\mathbb{R}^T$, which will be denoted by $\mathcal{R}^T$. By construction, all the functions Y_t are measurable with respect to $\mathcal{R}^T$. In what follows, we show that this σ-algebra is equal to the union over all countable (not necessarily finite !) families of indices $S \subset T$, S countable, of the σ-algebras generated by sets called *cylinders*. This shall have important consequences. Roughly, we can say that the σ-algebra $\mathcal{R}^T$ answers only "countable questions".

Consider the class $\mathcal{R}_0^T$ of the sets of the form :

$$\begin{aligned} A &= \{x \in \mathbb{R}^T : (Y_{t_1}(x), Y_{t_2}(x), ..., Y_{t_k}(x)) \in H\} \\ &= \{x \in \mathbb{R}^T : (x(t_1), x(t_2), ..., x(t_k)) \in H\} \end{aligned}$$

where k is some integer, $(t_1, t_2, ..., t_k)$ is a k-tuple of distinct points of T, and $H \in \mathcal{B}(\mathbb{R}^k)$. Sets of this form are called **cylinders**. Basically, the set A above is the set of functions that, evaluated at a *finite* set of points $(t_1, t_2, ..., t_k)$, have a certain property, in this case, belonging to a certain subset $H \in \mathcal{B}(\mathbb{R}^k)$. In particular, if $f \in A$, and g is a function that agrees with f on $(t_1, t_2, ..., t_k)$, then $g \in A$. If we fix $k = 1$, we see that $\mathcal{R}_0^T$ generate $\mathcal{R}^T$.

If A is as above, we have that $\mathbb{R}^T \setminus A = \{x \in \mathbb{R}^T : (x(t_1), x(t_2), ..., x(t_k)) \in \mathbb{R}^k \setminus H\}$. This means that $\mathcal{R}_0^T$ is closed under complementation. To show that $\mathcal{R}_0^T$ is closed under finite unions, we first re-write its sets is general form. Suppose A is given as above and B is given by :

$$\begin{aligned} B &= \{x \in \mathbb{R}^T : (Y_{s_1}(x), Y_{s_2}(x), ..., Y_{s_j}(x)) \in I\} \\ &= \{x \in \mathbb{R}^T : (x(s_1), x(s_2), ..., x(s_j)) \in I\} \end{aligned}$$

where j is some integer, $(s_1, s_2, ..., s_j)$ is a j-tuple of distinct points of T, and $I \in \mathcal{B}(\mathbb{R}^k)$. As in the previous part, we make use of a larger tuple $(u_1, u_2, ..., u_m)$ containing all the t_α and all the s_β. Again, $(t_1, t_2, ..., t_k)$ must be the initial segment of some permutation of $(u_1, u_2, ..., u_m)$ and if $\psi = \psi_t$ is as

before while ψ_s is the corresponding function where we consider $(s_1, s_2, ..., s_j)$ instead of $(t_1, t_2, ..., t_k)$, we let $H' = \psi_t^{-1}(H) \subset \mathbb{R}^m$ and $I' = \psi_s^{-1}(I) \subset \mathbb{R}^m$. Thus $H', I' \in \mathcal{B}(\mathbb{R}^m)$ and :

$$A = \{x \in \mathbb{R}^T : (x(u_1), x(u_2), .., x(u_m)) \in H'\} \tag{1.8}$$

while

$$B = \{x \in \mathbb{R}^T : (x(u_1), x(u_2), .., x(u_m)) \in I'\}. \tag{1.9}$$

Having said that, we almost finished the job, since we see that the expressions of A and B are very similar and we can write :

$$A \cup B = \{x \in \mathbb{R}^T : (x(u_1), x(u_2), .., x(u_m)) \in H' \cup I'\}.$$

which has the exact same form as the sets of $\mathcal{R}_0^T$; a union of cylinders is again a cylinder. We conclude that $\mathcal{R}_0^T$ is an algebra generating $\mathcal{R}^T$. Since all the mappings of $\{Y_t, t \in T\}$ are measurable with respect to $\mathcal{R}^T$, and since $(\mathbb{R}^T, \mathcal{R}^T)$ forms a measurable space, it remains to define a measure on $\mathcal{R}_0^T$ to get a probability space for which $(Y_t, t \in T)$ is be a stochastic process. The *Kolmogorov's Extension Theorem* allows us to do so.

1.3 Kolmogorov's Extension Theorem

In the previous section, we have constructed a measurable space $(\mathbb{R}^T, \mathcal{R}^T)$. We now investigate the options available to define a probability measure P on this space in order to make it a probability space. A consequence of this is that the Y_t will become random variables on $(\mathbb{R}^T, \mathcal{R}^T, P)$. In the case where $T = \mathbb{R}_+$, $(Y_t, t \in \mathbb{R}_+)$ will be a stochastic process on $(\mathbb{R}^{\mathbb{R}_+}, \mathcal{R}^{\mathbb{R}_+}, P)$ and, in the next section, we will have to discuss the definition of its distribution.

Theorem 3 (Kolmogorov's Extension Theorem)**.** *If $\mu_{t_1 t_2,..,t_k}$ are a system of distribution satisfying the consistency conditions* (1.2) *and* (1.3)*, then there exist a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process $(X_t, t \in T)$ on this space having $\mu_{t_1 t_2,..,t_k}$ as its finite dimensional distributions.*

We see that this theorem can be reformulated as follows :

Theorem 4 (Reformulation of Kolmogorov's Extension Theorem)**.** *If $\mu_{t_1 t_2,..,t_k}$ are a system of distribution satisfying the consistency conditions* (1.2) *and* (1.3)*, then there is a probability measure P on $\mathcal{R}^T$ such that the coordinate-mappings process $(Y_t, t \in T)$ on $(\mathbb{R}^T, \mathcal{R}^T, P)$, has $\mu_{t_1 t_2,..,t_k}$ as its finite dimensional distributions.*

Indeed, it is clear that the latter implies the former. Now, we investigate the converse. We want to show that if $(X_t, t \in T)$ is as in the Kolmogorov's extension theorem then there is a probability measure P on $\mathcal{R}^T$ such that the coordinate-mappings process $(Y_t, t \in T)$ on $(\mathbb{R}^T, \mathcal{R}^T, P)$, has $\mu_{t_1 t_2,..,t_k}$ as its finite dimensional distributions.

If $(X_t, t \in T)$ is a stochastic process on $(\Omega, \mathcal{F}, P)$ having $\mu_{t_1 t_2,..,t_k}$ as its finite dimensional distributions, then we can define a map $\gamma : \Omega \longrightarrow \mathbb{R}^T$ as :

$$\gamma(\omega) = Y_t^{-1}(X_t(\omega)), \qquad t \in T. \tag{1.10}$$

We see that γ satisfies $Y_t(\gamma(\omega)) = X_t(\omega)$. Equality (1.10) shows that, for each $\omega \in \Omega$, $\gamma(\omega)$ is in $\mathbb{R}^T$ and that it is a function equal to $X_t(\omega)$ at point t, for any $t \in T$.Also :

$$\begin{aligned} &\gamma^{-1}\{x \in \mathbb{R}^T : (x(t_1), x(t_2,), ..., x(t_k)) \in H\} \\ &= \{\omega \in \Omega : (Y_{t_1}(\gamma(\omega)), Y_{t_2}(\gamma(\omega)), ..., Y_{t_k}(\gamma(\omega))) \in H\} \\ &= \{\omega \in \Omega : (X_{t_1}(\omega), X_{t_2}(\omega), ..., X_{t_k}(\omega)) \in H\} \end{aligned} \tag{1.11}$$

This set is in $\mathcal{F}$ if $H \in \mathcal{B}(\mathbb{R}^k)$ since $X_{t_1}, X_{t_2}, ..., X_{t_k}$ are random variables (measurable functions from Ω to $\mathbb{R}$). Thus, $\gamma^{-1}(A) \in \mathcal{F}$ if A is in $\mathcal{R}_0^T$. So we have that $\gamma : \Omega \longrightarrow \mathbb{R}^T$ is a measurable function. By assumption :

$$\begin{aligned} P(\gamma^{-1}\{x \in \mathbb{R}^T : (Y_{t_1}(x), Y_{t_2}(x), ..., Y_{t_k}(x)) \in H\}) &= P(\gamma^{-1}\{x \in \mathbb{R}^T : (x(t_1), x(t_2,), ..., x(t_k)) \in H\}) \\ &= P(\omega \in \Omega : (X_{t_1}(\omega), X_{t_2}(\omega), ..., X_{t_k}(\omega)) \in H) \\ &= \mu_{t_1 t_2,..,t_k}(H) \end{aligned}$$

Thus, the coordinate mappings process $(Y_t, t \in T)$ on $(\mathbb{R}^T, \mathcal{R}^T, P\gamma^{-1})$ also has the finite dimensional distributions $\mu_{t_1 t_2,..,t_k}(H)$ even though it is defined on a different probability space than X ! So the two formulations of the Kolmogorov's extension theorem are strictly equivalent.

Proof of the Kolmogorov's extension theorem. See [5], p489-492. □

Theorem 5. *Let $(X_t, t \in T)$ be a family of real function on Ω.*

1. *If $A \in \sigma(X_t, t \in T)$, $\omega \in A$, and if $X_t(\omega) = X_t(\omega')$ for all $t \in T$, then $\omega' \in A$.*
2. *If $A \in \sigma(X_t, t \in T)$, then $A \in \sigma(X_t, t \in S)$ for some countable subset S of T.*

Proof. We start with a preliminary step. As before, let $\gamma : \Omega \longrightarrow \mathbb{R}^T$ be so that :

$$\gamma(\omega) = Y_t^{-1}(X_t(\omega)), \qquad t \in T. \tag{1.12}$$

or $Y_t(\gamma(\omega)) = X_t(\omega)$. If we set $\mathcal{F} = \sigma(X_t, t \in T)$, we can use the previous results (specifically (1.11)) to state that γ is measurable with respect to $\mathcal{F}$. And thus, $\mathcal{F}$ must contain the class $\{\gamma^{-1}(B)$, where $B \in \mathcal{R}^T\}$, which is a σ-field. By (1.11), this class contains the sets $\{\omega \in \Omega : (X_{t_1}(\omega), X_{t_2}(\omega), ..., X_{t_k}(\omega)) \in H\}$, $H \in \mathcal{B}(\mathbb{R}^k)$, and thus the σ-field $\mathcal{F}$ they generated. So the double inclusion gives :

$$\sigma(X_t, t \in T) = \{\gamma^{-1}(B), \text{ where } B \in \mathcal{R}^T\}$$

We can now prove the theorem. The first hypothesis gives $\omega \in A = \gamma^{-1}(B)$, for some $B \in \mathcal{R}^T$. Also, the hypotheses imply that $\gamma(\omega) = Y_t^{-1}(X_t(\omega)) = Y_t^{-1}(X_t(\omega')) = \gamma(\omega')$, so we necessarily have $\omega' \in A$. That's for the first point. Now, for $S \subset T$, S countable, let $\mathcal{F}_S = \sigma(X_t, t \in S) \subset \mathcal{F}$. We want to show $\mathcal{F} = \cup_S \mathcal{F}_S$ where the union is taken over all countable subsets S of T. If $A_1, A_2, ..$ are in $\bigcup_S \mathcal{F}_S$, then $A_n \in \mathcal{F}_{S_n}$ for some countable subsets S_n of T. And since $\mathcal{F}_{S_n} \subset \mathcal{F}_{\cup_m S_m}$ $(A_n \in \mathcal{F}_{\cup_m S_m})$, we also have :

$$\bigcup_n A_n \subset \mathcal{F}_{\cup_n S_n} \subset \bigcup_S \mathcal{F}_S$$

which implies that $\bigcup_S \mathcal{F}_S$ is a σ-field. And since it contains the sets $\{\omega \in \Omega : X_t(\omega) \in H\}$ (recall $\mathcal{F}_S = \sigma(X_t, t \in S)$), it contains the σ-field $\mathcal{F}$ they generate : $\mathcal{F} \subset \bigcup_S \mathcal{F}_S$. Thus $\mathcal{F} = \cup_s \mathcal{F}_S$. □

the previous theorem tells us that :

$$\sigma(X_t, t \in T) = \bigcup_{\substack{S \subset T \\ S\text{countable}}} \sigma(X_{t_i}^{-1}(B), B \in \mathcal{B}(\mathbb{R}), t_i \in S). \tag{1.13}$$

In the case of the product space $\mathbb{R}^{\mathbb{R}_+}$, we have that :

$$\mathcal{R}^{\mathbb{R}_+} = \sigma(Y_t, t \in T) = \bigcup_{\substack{S \subset T \\ S\text{countable}}} \sigma(Y_{t_i}^{-1}(B), B \in \mathcal{B}(\mathbb{R}), t_i \in S). \tag{1.14}$$

We now state an important consequence of this property.

Theorem 6. *Let $T = \mathbb{R}_+$, and so $\mathbb{R}^T = \mathbb{R}^{\mathbb{R}_+}$. Let $\mathcal{C} = \mathcal{C}(\mathbb{R}_+, \mathbb{R})$ be the subset of $\mathbb{R}^{\mathbb{R}_+}$ which contains only the continuous functions from $\mathbb{R}_+$ to $\mathbb{R}$. Then $\mathcal{C}$ is not measurable with respect to $\mathcal{R}^{\mathbb{R}_+}$*

Proof. Suppose $\mathcal{C}$ is measurable with respect to $\mathcal{R}^{\mathbb{R}_+}$, that is $\mathcal{C} \in \bigcup_{\substack{S \subset T \\ S\text{countable}}} \sigma(Y_{t_i}^{-1}(B), B \in \mathcal{B}(\mathbb{R}), t_i \in S)$. We must have that $\mathcal{C} = \sigma(Y_t, t \in S)$ for some countable subset S of T. Using now the first part of the previous theorem, we have that if $f \in \mathcal{C}$ and if $g \in \mathbb{R}^{\mathbb{R}_+}$ is a function so that $f(t) = g(t)$ for all $t \in S$, then $g \in \mathcal{C}$. But this is absurd, whatever this S might be. □

1.4 Law of a Stochastic Process

In this section, we define precisely what we mean by the "law" of a stochastic process X. We use the same notations as before. Recall that the functions $Y_t : (\mathbb{R}^{\mathbb{R}_+}, \mathcal{R}^{\mathbb{R}_+}) \longmapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\omega \longmapsto \omega(t)$ are the coordinate mappings. If $(X_t, t \in \mathbb{R}_+)$ is a stochastic process on $(\Omega, \mathcal{F}, P)$ taking its values in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the application $\phi : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\omega \longmapsto \phi(\omega) \equiv (\omega : t \longmapsto X_t(\omega))$ is measurable since $X_t = Y_t \circ \phi$ is measurable for all $t \in \mathbb{R}_+$. So in overall, here are the applications we are dealing with :

$$\begin{aligned} X_t \colon (\Omega, \mathcal{F}, P) &\xrightarrow{\quad\phi\quad} (\mathbb{R}^{\mathbb{R}_+}, \mathcal{R}^{\mathbb{R}_+}) \xrightarrow{\quad Y_t \quad} (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ \omega &\longmapsto (\omega : t \longmapsto X_t(\omega)) \longmapsto \omega(t) = X_t(\omega). \end{aligned}$$

Following the definition of the first section, we can define the image-measure $\phi(P)$ of the measure P by the application $\phi : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}^{\mathbb{R}_+}, \mathcal{R}^{\mathbb{R}_+})$, that is the measure P_X on $\mathcal{R}^{\mathbb{R}_+}$ so that :

$$P_X(A) = P(\phi^{-1}(A))$$

for all $A \in \mathcal{R}^{\mathbb{R}_+}$. In particular if $t_1 \leq t_2 \leq ... \leq t_k$ is a finite subset of $\mathbb{R}_+$, and $A_i \in \mathcal{B}(\mathbb{R})$ then :

$$\begin{aligned} &P(\omega \in \Omega : X_{t_1}(\omega) \in A_1, X_{t_2}(\omega) \in A_2, ..., X_{t_k}(\omega) \in A_k) \\ &= P(\omega \in \Omega : (Y_{t_1} \circ \phi)(\omega) \in A_1, (Y_{t_2} \circ \phi)(\omega) \in A_2, ..., (Y_{t_k} \circ \phi)(\omega) \in A_k) \\ &= P(\omega \in \Omega : \omega \in \bigcap_{i=1}^{k} \phi^{-1}(Y_{t_i}^{-1}(A_i))) \\ &= P(\phi^{-1}(\bigcap_{i=1}^{k} Y_{t_i}^{-1}(A_i))) \\ &= P_X(\bigcap_{i=1}^{k} Y_{t_i}^{-1}(A_i)) = P_X(\omega \in \mathbb{R}^{\mathbb{R}_+} : Y_{t_1}(\omega) \in A_1, Y_{t_2}(\omega) \in A_2, ..., Y_{t_k}(\omega) \in A_k) \end{aligned}$$

Or, more simply :

$$P(X_{t_1} \in A_1, X_{t_2} \in A_2, ..., X_{t_k} \in A_k) = P_X(Y_{t_1} \in A_1, Y_{t_2} \in A_2, ..., Y_{t_k} \in A_k). \tag{1.15}$$

We say that X and Y are **versions** of each other [1]. The process Y is called the **canonical version** of the process X. And the probability P_X is called the **law** of the process X.

1.5 Finite-Dimensional Distributions Determine the Law of a Process

The section on finite distributions allowed us to partially characterize stochastic processes. Why only partially ? Well, since we were only concerned with *finite* sequences of times $t_1, t_2, ..., t_k$, and the associated joint distibutions, we can hardly say anything about the global properties of sample paths $t \longmapsto X_t(\omega)$ or asymptotics. Now, even if some properties cannot be deduced from finite dimensional distributions, these distributions define "uniquely" the infinite dimensional distribution. That is what the next theorem tells us.

Theorem 7. *Let $X = (X_t, t \in T)$ and $Y = (Y_t, t \in T)$ be two real valued processes on T. Then X and Y have the same distribution (or law) if and only if their finite-dimensional distributions agree.*

Proof. See [6], p16. □

[1] More generally, we say that two processes are versions of each other if all they finite dimensional distributions agree

Chapter 2

Continuity of Stochastic Processes

In this chapter, we give some introductory definitions (modifications, indistinguishability) and we explain why continuity of the trajectories of a stochatic process is important. Finally, we state the very famous Kolmogorov continuity criterion and give a detail proof of this result. This step is fundamental since it will allow us to construct a realistic Brownian motion.

2.1 Similarites Between Processes

Two processes $X = (X_t, t \in T)$ and $Y = (Y_t, t \in T)$ defined on the same probability space $(\Omega, \mathcal{F}, P)$ are said to be **modifications** of each other if :

$$\text{for each fixed } t_0 \in T \qquad P(\omega \in \Omega : X_{t_0}(\omega) = Y_{t_0}(\omega)) = 1$$

that is, for each fixed t_0, the positions of X and Y at time t_0 are the same almost surely.

Two processes $X = (X_t, t \in T)$ and $Y = (Y_t, t \in T)$ defined on the same probability space $(\Omega, \mathcal{F}, P)$ are said to be **indistinguishable** if

$$P(\omega \in \Omega : \text{for each fixed } t_0 \in T,\ X_{t_0}(\omega) = Y_{t_0}(\omega)) = 1$$

that is, almost surely, they have the same trajectories. In this case, we don't have to fix t_0 in advance ; the expression "for each fixed $t_0 \in T$" enters the definition of the event which has probability 1 while in the previous case, we had to fix t_0 before taking the probability. And since for each fixed $t_0 \in T$: $\{\omega \in \Omega : X_{t_0}(\omega) = Y_{t_0}(\omega)\} \supset \{\omega \in \Omega : \text{for each fixed } t_0' \in T,\ X_{t_0'}(\omega) = Y_{t_0'}(\omega)\}$, we necessarily have that if two processes are indistinguishable, then they are modifications of each other. So the indistinguishability property is stronger than the "modification property". The latter gives us information about fixed points of the trajectories (of X and Y) while the former characterise the trajectories as a whole.

2.2 Continuity of Trajectories

As said above, our goal is to construct rigorously a probabilty space and a stochastic process that will be a model for a particle moving randomly in space. This means that we need to fulfil some requirements. In particular, it is very important for the trajectories of the stochastic processes considered to be continuous. Here, we do not discuss the issue of continuity of space time and we will assume it is a continuum.

The process $X = (X_t, t \in T)$ is said to have, almost surely, **continuous trajectories** if it is such that :

$$P(\omega \in \Omega : t \longmapsto X_t(\omega) \text{ is continuous over } T\ ) = 1.$$

If $\mathcal{N}$ is the set (of probability 0) where X doesn't have continuous trajectories, we can set, $\forall \omega \in \mathcal{N}$, $X_t(\omega) = 0$ for all $t \in T$ and we obtain a stochastic process with continuous trajectories (everywhere on

Ω) with the same law as before.

Having defined the notion of continuity for the trajectories of a stochastic process, we can give another argument justifying the need for this property.

Theorem 8 (A good reason to have continuous trajectories)**.** *Let $X = (X_t, t \in \mathbb{R}_+)$ be some real stochastic process on Ω with continuous trajectories. Then, on any bounded interval I, $M : \Omega \longrightarrow \mathbb{R}$, $\omega \longmapsto M(\omega) = \sup_{t\in I} X_t(\omega)$ is a random variable. In particular, it is measurable.*

Proof. First, we have that $M(\omega)$ must be finite on I since this interval is bounded and $t \longmapsto X_t(\omega)$ is continuous *everywhere* on $\mathbb{R}_+$. Then, we notice that $M(\omega) > a$ if and only if $X_t(\omega) > a$ for some $t \in I$. Let's fix any of these "t"s and call it t_0. By density of $I' = I \cap \mathbb{Q} \subset I$ in I, there is some sequence of rationals (q_n) in I' so that $\lim q_n = t_0$. By *continuity* of $t \longmapsto X_t(\omega)$ we must have the existence of $N \in \mathbb{N}$ so that $X_{q_n}(\omega) > a$ for $n > N$. Actually, we can write :

$$M^{-1}((a, \infty)) = \{\omega \in \Omega : M(\omega) > a\} = \bigcup_{q\in I'} \{\omega \in \Omega : X_q(\omega) > a\}.$$

Since all the sets of the union are measurable (because for all $t \in \mathbb{R}_+$ X_t is a random variable), the union itself is measurable. So by definition, $M : \Omega \longrightarrow \mathbb{R}$ is a measurable function on a probability space, that is, it is a random variable. To conclude we observe that if $M : \Omega \longrightarrow \mathbb{R}$, $\omega \longmapsto M(\omega) = \sup_{t\in I} X_t(\omega)$ is measurable, then $M' : \Omega \longrightarrow \mathbb{R}$, $\omega \longmapsto M'(\omega) = \sup_{t\in I} -X_t(\omega)$ is also measurable as well as $m : \Omega \longrightarrow \mathbb{R}$, $\omega \longmapsto -M'(\omega) = \inf_{t\in I} X_t(\omega)$. □

It is important to realise that the measurability of the random variable M above is not unimportant. Indeed, controlling the finite dimensional distributions of a process is not enough to guarantee that it has measurable extrema on bounded intervals. And if, for example, we construct a stochastic model for some "real world" application, say to approximate prices on a market, it could be very embarassing not to be able to make statements about the supremum of the process on some bounded interval.

Here is another good consequence of continuity. We said above that indistinguishability and modification properties were different for a given couple of stochastic processes. However, it is not true anymore if X and Y (almost surely) have continuous trajectories. In this case, if we state that, for each fixed $t_0 \in T$, the positions of X and Y at time t_0 are almost surely the same, it is equivalent to saying that X and Y almost surely have the same trajectories. So the statement about the relative positions at any fixed time becomes a statement about the relative positions at all times. To prove that, we use a density argument.

Let $T = \mathbb{R}_+$ represent time [1]. We restrict $T' = T \cap \mathbb{Q}$ (T' is clearly dense in T) to see that :

$$\{\omega \in \Omega : \text{for each fixed } q_0 \in T',\ X_{q_0}(\omega) = Y_{q_0}(\omega)\} = \bigcap_{q_0\in T'} \{\omega \in \Omega : X_{q_0}(\omega) = Y_{q_0}(\omega)\}$$

so, if for each fixed $q_0 \in T'$, we have $P(\omega \in \Omega : X_{q_0}(\omega) = Y_{q_0}(\omega)) = 1$, then we must have $P(\omega \in \Omega : \text{for each fixed } q_0 \in T',\ X_{q_0}(\omega) = Y_{q_0}(\omega)) = 1$. Now, we claim that for processes X and Y having (almost surely) continuous trajectories, $\{\omega \in \Omega : \text{for each fixed } q_0 \in T',\ X_{q_0}(\omega) = Y_{q_0}(\omega)\} = \{\omega \in \Omega : \text{for each fixed } t_0 \in T,\ X_{t_0}(\omega) = Y_{t_0}(\omega)\}$. Indeed, suppose ω_0 is in the first set, but not the second. Then we have that for each $q_0 \in T'$, $X_{q_0}(\omega_0) = Y_{q_0}(\omega_0)$, but there exists $t_0 \in T$ so that $X_{t_0}(\omega_0) \neq Y_{t_0}(\omega_0)$. Now, by density of T' in T, we can find a sequence (q_n) of points in T' so that $\lim q_n = t_0$. Since for all $n \in \mathbb{N}$, $q_n \in T'$, we have : $X_{q_n}(\omega_0) = Y_{q_n}(\omega_0)$ for all $n \in \mathbb{N}$. Taking the limit on both sides of the equation and using the continuits of both X and Y, we have that : $X_{t_0}(\omega_0) = Y_{t_0}(\omega_0)$. Which is a contradiction with the previous assumption. Thus, there is no such ω_0. This shows one inclusion and since $T' \subset T$, the other is trivial.

Having discussed the preceding issues, we understood that the continuity of the trajectories of a stochastic process can be very useful. But we still don't know which processes have this property or if they don't, which processes can be "transformed" to acquire it. The goal of the next section is to answer this question.

[1] Here, we consider $T = \mathbb{R}_+$ for simplicity, but the proof can be generalised

2.3 Kolmogorov's Criterion for Continuity

Theorem 9 (Kolmogorov's Criterion for Continuity)**.** *Let $X = (X_t, t \in [0,1])$ be a real-valued process. Suppose that there exist $(\gamma, C, \epsilon) \in (\mathbb{R}_+^*)$ so that :*

$$\forall s,t \in [0,1], \qquad \mathrm{E}(|X_t - X_s|^\gamma) \leq C \cdot |t-s|^{1+\epsilon}. \tag{2.1}$$

Then, there exists a modification $Y = (Y_t, t \in [0,1])$ of X so that, if we define $S_\alpha : \Omega \longrightarrow \mathbb{R}$, $\omega \longmapsto \sup\{\frac{|Y_t(\omega)-Y_s(\omega)|}{|t-s|^\alpha} : \quad s,t \in [0,1],\ s \neq t\}$, we have :

- *If $\gamma \geq 1$*

$$\forall \alpha \in [0, \frac{\epsilon}{\gamma}), \qquad \mathrm{E}(S_\alpha^\gamma) < \infty.$$

- *If $\gamma < 1$*

$$\forall \alpha \in [0, \epsilon), \qquad \mathrm{E}(S_\alpha^\gamma) < \infty.$$

From measure theory, we have that if $\mathrm{E}(S_\alpha^\gamma) < \infty$, then $P(\omega \in \Omega : S_\alpha^\gamma(\omega) < \infty) = 1$. And so :

$$P(\omega \in \Omega : \sup\{\frac{|Y_t(\omega) - Y_s(\omega)|}{|t-s|^\alpha} : \quad s,t \in [0,1],\ s \neq t\} < \infty) = 1$$

which is equivalent to saying that the trajectory of Y is almost surely Hölder continuous of order α. In particular, it is continuous (almost surely).

Proof. The proof of the theorem contains six steps.

Step 0 : we review some results on dyadic numbers

Recall that for all $x \in [0,1)$, there exists a sequence (x_k) of numbers in $\{0,1\}$ so that $x = \sum_{k=1}^\infty \frac{x_k}{2^k}$. If this series is finite (let's say that it contains m terms), x is called a *dyadic number* and $2^m x \in \mathbb{N}$ so that $x = \frac{k}{2^m}$ for some $k \in \mathbb{N}$ such that $k \leq 2^{m-1}x_1 + ... + 2x_{m-1} + x_m \leq 2^m - 1$. The two representations above (as a sum or a ratio) are equal in the sense that both can represent all the dyadic numbers of $[0,1)$. Also, it is straight forward that the dyadic numbers are dense in $[0,1]$.
Now, if $D_m = \{\frac{k}{2^m} : k \in \{0,1,...,2^m - 1\}\}$ then the set D of all the dyadic numbers of $[0,1)$ is equal to $\bigcup_{m\in\mathbb{N}} D_m$ and we have $D_m \subset D_{m+1}$. If $d \in D$ is a dyadic number, and if we choose its representation as a finite series ($d = \sum_{k=1}^m \frac{\delta_k}{2^k}$ where $\delta_i \in \{0,1\}$), we can consider its partial sums $d_n = \sum_{k=1}^n \frac{\delta_k}{2^k}$ to state that there exists at least one increasing sequence (d_n) of dyadic numbers so that $d_n \in D_n$ for all $n \in \mathbb{N}$ and $d_n = d$ for n large enough. By construction, the difference $d_{n+1} - d_n$ can either be 0 or $\frac{1}{2^{n+1}}$ (when we consider the sequence constructed from partial sums).
For the same reason, we can find, for each $x \in [0,1]$, an increasing sequence (d_n) of dyadic numbers so that $\lim d_n = x$ and $d_n \in D_n$ for all $n \in \mathbb{N}$. With the same notations as before, it suffices to take $d_n = \sum_{k=1}^n \frac{x_k}{2^k}$ and we have $|d_n - x| \leq \frac{1}{2^n}$ for all $n \in \mathbb{N}$. Again, by construction, the difference $d_{n+1} - d_n$ can either be 0 or $\frac{1}{2^{n+1}}$. In this case however, we do not necessarily have the property that $d_n = d$ for n large enough.
Let's fix $s,t \in [0,1)$ and (s_n), (t_n) two sequences (as above) converging to s and t respectively. Since we necessarily have $|s-t| \leq \frac{1}{2^0}$ and $\lim \frac{1}{2^m} = 0$, there must exist an m, maximum for this property, so that $|s-t| \leq \frac{1}{2^m}$. Maximality of m implies, in particular, $\frac{1}{2^{m+1}} < |t-s| \leq \frac{1}{2^m}$. Buy re-writing $[0,1) = \bigcup_{k=0}^{2^m-1} [\frac{k}{2^m}, \frac{k+1}{2^m})$, two cases occur. Either s,t belong to the same interval, or they don't. If they do, there exists a $k \in \{0,1,...,2^m-1\}$ so that $s,t \in [\frac{k}{2^m}, \frac{k+1}{2^m})$. This forces $s_m = t_m = \frac{k}{2^m}$. If s,t do not belong to the same interval, then there exists a $k \in \{0,1,...,2^m-1\}$ so that (without loss of generality, we suppose $s < t$) :

$$s \in [\frac{k-1}{2^m}, \frac{k}{2^m}) \qquad \text{and} \qquad t \in [\frac{k}{2^m}, \frac{k+1}{2^m})$$

which implies $s_m = \frac{k-1}{2^m}$ and $t_m = \frac{k}{2^m}$ and in particular $|s_m - t_m| = \frac{1}{2^m}$.

Step 1 : we define the terms

We define Δ_m, a subset of the dyadic numbers of $[0,1)$ by :

$$\Delta_m = \{(s,t) \in D^2 : |s-t| = \frac{1}{2^m}\}$$

Let's estimate the cardinal of this set. If $(s,t) \in \Delta_m$, there are 2^m choices for s (precisely : $0/2^m, 1/2^m, ..., (2^m-1)/2^m$). Now if $s \neq 0, s \neq (2^m-1)/2^m$, there are $2^m - 2$ choices for s and two choices (in each case) are left for t so $(2^m-2)2$ in total. If $s = 0$ or $s = (2^m-1)/2^m$, we have only one choice (in each case) left for t. So in total, there are $(2^m-2)2 + 2 = 2(2^m-1) \leq 2^{m+1}$ choices. We thus have $|\Delta_m| \leq 2^{m+1}$.

Let's fix $m \in \mathbb{N}$. We define

$$\begin{aligned} K_m \colon \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto K_m(\omega) = \sup\{|X_t(\omega) - X_s(\omega)|, (s,t) \in \Delta_m\}, \end{aligned}$$

note that the supremum is taken over a finite set.
Let's fix $\alpha \geq 0$. We define

$$\begin{aligned} M_\alpha \colon \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto M_\alpha(\omega) = \sup\{\frac{|X_t(\omega) - X_s(\omega)|}{|t-s|^\alpha},\ s,t \in D\ s \neq t\}, \end{aligned}$$

Step 2 : we find an upper bound for $\mathrm{E}(K_m^\gamma)$

$$\begin{aligned} \mathrm{E}(K_m^\gamma) &= \mathrm{E}((\sup\{|X_t - X_s|, (s,t) \in \Delta_m\})^\gamma) \\ &\leq \mathrm{E}(\sum_{(s,t)\in\Delta_m} |X_t - X_s|^\gamma) \\ &\overset{|\Delta_m|\leq\infty}{=} \sum_{(s,t)\in\Delta_m} \mathrm{E}(|X_t - X_s|^\gamma) \\ &\overset{\text{hyp}}{\leq} \sum_{(s,t)\in\Delta_m} C \cdot |t-s|^{1+\epsilon} \\ &\leq C|\Delta_m| \max\{|t-s|^{1+\epsilon}, (s,t) \in \Delta_m\} \\ &\leq C2^{m+1}(2^{-m})^{1+\epsilon} = C2^{m+1-m(1+\epsilon)} = J2^{-m\epsilon} \text{ where } J \in \mathbb{R}. \end{aligned}$$

Step 3 : we find an upper bound $|X_t - X_s|$

Let $s,t \in D$ satisfy $|t-s| \leq \frac{1}{2^m}$. We consider two increasing sequences of dyadic numbers (as above) $(s_n), (t_n)$ converging to s and t respectively. In step 1, we showed that there exist integers n_0 and n_1 so that $s_i = s$ and $t_j = t$ for $i > n_0$ and $j > n_1$. In particular :

$$\begin{aligned} X_{s_{i+1}} - X_{s_i} &= 0 \quad \text{for } i > n_0 \\ X_{t_i} - X_{t_{i+1}} &= 0 \quad \text{for } i > n_1. \end{aligned}$$

We thus have (recall that $|s_{i+1} - s_i| \in \{0, \frac{1}{2^{m+1}}\}$ and the same for (t_n)) :

$$\begin{aligned} |X_s - X_t| &= |X_{s_m} + (X_{s_{m+1}} - X_{s_m}) + (X_{s_{m+2}} - X_{s_{m+1}}) + ... + (X_{s_{n_0+1}} - X_{s_{n_0}}) + 0 + 0 + ... \\ &\quad - X_{t_m} + (X_{t_m} - X_{t_{m+1}}) + (X_{t_{m+1}} - X_{t_{m+2}}) + ... + (X_{t_{n_1}} - X_{t_{n_1+1}}) + 0 + 0 + ...| \\ &= |\sum_{i=m}^{\infty}(X_{s_{i+1}} - X_{s_i}) + X_{s_m} - X_{t_m} + \sum_{i=m}^{\infty}(X_{t_i} - X_{t_{i+1}})| \\ &\overset{\text{step 3}}{\leq} K_m + 2\sum_{i=m}^{\infty} K_{i+1} \\ &= K_m + 2\sum_{i=m+1}^{\infty} K_i \leq 2\sum_{i=m}^{\infty} K_i. \end{aligned}$$

Step 4 : we find an upper bound for M_α and its γth moment

Our goal is to find an upper bound for M_α, and this random variable is the supremum of some random set. So, what we need to do is to find an upper bound for the supremum of this set. To do so, we construct a new (simpler) random set, and show that its supremum is greater that the supremum of the "old" set. More precisely, we want an upper bound for $\sup\{\frac{|X_t-X_s|}{|t-s|^\alpha} : \ s,t \in D \ s \neq t\}$. Let pick any point in this set, it is of the form $\frac{|X_t-X_s|}{|t-s|^\alpha}$ for some *fixed* $s,t \in D \ s \neq t$. Again, there must be some m, maximum for this property, such that : $|t-s| \leq 1/2^m$. In particular :

$$
\begin{aligned}
|\frac{1}{2^{m+1}}| < |t-s| \leq |\frac{1}{2^m}| &\Longrightarrow |\frac{1}{2^{m+1}}|^\alpha < |t-s|^\alpha \leq |\frac{1}{2^m}|^\alpha \\
&\Longrightarrow \frac{1}{|\frac{1}{2^{m+1}}|^\alpha} > \frac{1}{|t-s|^\alpha} \geq \frac{1}{|\frac{1}{2^m}|^\alpha} \\
&\Longrightarrow \frac{|X_t-X_s|}{|\frac{1}{2^{m+1}}|^\alpha} \geq \frac{|X_t-X_s|}{|t-s|^\alpha} \geq \frac{|X_t-X_s|}{|\frac{1}{2^m}|^\alpha} \\
&\overset{|t-s|\leq\frac{1}{2^m}}{\Longrightarrow} \sup\{\frac{|X_{t'}-X_{s'}|}{|\frac{1}{2^{m+1}}|^\alpha} : |t'-s'| \leq \frac{1}{2^m}\} \geq \frac{|X_t-X_s|}{|t-s|^\alpha}
\end{aligned}
$$

What we showed here is that, for any point in $\{\frac{|X_t-X_s|}{|t-s|^\alpha}, \ s,t \in D \ s \neq t\}$, we can find a point in $\{\sup\{\frac{|X_t-X_s|}{|\frac{1}{2^{m+1}}|^\alpha}, |t-s| \leq \frac{1}{2^m}\} : m \in \mathbb{N}, s,t \in D \ s \neq t\}$ that is larger. So we must have :

$$
\begin{aligned}
M_\alpha &= \sup\{\frac{|X_t-X_s|}{|t-s|^\alpha} : \ s,t \in D \ s \neq t\} \\
&\leq \sup\{\sup\{\frac{|X_t-X_s|}{|\frac{1}{2^{m+1}}|^\alpha}, |t-s| \leq \frac{1}{2^m}\} : m \in \mathbb{N}, s,t \in D \ s \neq t\}
\end{aligned}
$$

which implies

$$
\begin{aligned}
M_\alpha &\leq \sup\{\sup\{2^{(m+1)\alpha}|X_t-X_s|, |t-s| \leq \frac{1}{2^m}\} : m \in \mathbb{N}, s,t \in D \ s \neq t\} \\
&= \sup\{2^{(m+1)\alpha} \sup\{|X_t-X_s|, |t-s| \leq \frac{1}{2^m}\} : m \in \mathbb{N}, s,t \in D \ s \neq t\} \\
&\overset{\text{step 3}}{\leq} \sup\{2 \cdot 2^{(m+1)\alpha} \sup\{\sum_{i=m}^{\infty} K_i, |t-s| \leq \frac{1}{2^m}\} : m \in \mathbb{N}, s,t \in D \ s \neq t\} \\
&= \sup\{2 \cdot 2^{(m+1)\alpha} \sum_{i=m}^{\infty} K_i : m \in \mathbb{N}\} \\
&= 2^{(\alpha+1)} \sup\{2^{m\alpha} \sum_{i=m}^{\infty} K_i : m \in \mathbb{N}\} \\
&\overset{\text{sum starts at } m}{\leq} 2^{(\alpha+1)} \sup\{\sum_{i=m}^{\infty} K_i 2^{i\alpha} : m \in \mathbb{N}\} \\
&\leq 2^{(\alpha+1)} \sum_{i=0}^{\infty} K_i 2^{i\alpha} \qquad \text{or} \qquad M_\alpha \leq 2^{(\alpha+1)} \sum_{i=0}^{\infty} K_i 2^{i\alpha}.
\end{aligned}
$$

So we did find an upper bound for M_α. But the theorem makes a statement about $\mathrm{E}(S_\alpha^\gamma)$ and not only $\mathrm{E}(S_\alpha)$ so we also need to find an upper bound for $\mathrm{E}(M_\alpha^\gamma) = ||M_\alpha||_\gamma^\gamma$.

- Case 1 : $\gamma \geq 1$

$$\begin{aligned}
||M_\alpha||_\gamma &\leq ||2^{(\alpha+1)} \sum_{i=0}^{\infty} K_i 2^{i\alpha}||_\gamma \\
&\leq 2^{(\alpha+1)} \sum_{i=0}^{\infty} 2^{i\alpha} ||K_i||_\gamma \\
&\overset{\text{step 3}}{\leq} 2^{(\alpha+1)} \sum_{i=0}^{\infty} 2^{i\alpha} (J 2^{-i\epsilon})^{\frac{1}{\gamma}} \\
&= 2^{(\alpha+1)} J^{\frac{1}{\gamma}} \sum_{i=0}^{\infty} 2^{i\alpha} 2^{-i\epsilon\frac{1}{\gamma}} \\
&= 2^{(\alpha+1)} J^{\frac{1}{\gamma}} \sum_{i=0}^{\infty} 2^{i(\alpha-\frac{\epsilon}{\gamma})}
\end{aligned}$$

and for this series to converge, we must have $\alpha - \frac{\epsilon}{\gamma} < 0$ or $\alpha \in [0, \frac{\epsilon}{\gamma})$. In this case only, we can write :

$$[\mathrm{E}(M_\alpha^\gamma)]^{\frac{1}{\gamma}} = ||M_\alpha||_\gamma < \infty$$

- Case 2 : $\gamma < 1$

$\gamma < 1 \Longrightarrow x \longmapsto x^\gamma$ concave and by Jensen's inequality (ϕ concave $\Longrightarrow \mathrm{E}(\phi(X)) \leq \phi(\mathrm{E}(X))$, we have :

$$\begin{aligned}
\mathrm{E}(M_\alpha^\gamma) &\leq [\mathrm{E}(M_\alpha)]^\gamma \\
&\leq [\mathrm{E}(2^{(\alpha+1)} \sum_{i=0}^{\infty} K_i 2^{i\alpha})]^\gamma \\
&\leq [2^{(\alpha+1)} \sum_{i=0}^{\infty} 2^{i\alpha} \mathrm{E}(K_i)]^\gamma \\
&\leq [2^{(\alpha+1)} \sum_{i=0}^{\infty} 2^{i\alpha} (J 2^{-i\epsilon})]^\gamma \\
&= [2^{(\alpha+1)} J \sum_{i=0}^{\infty} 2^{i(\alpha-\epsilon)}]^\gamma
\end{aligned}$$

and for this series to converge we must have $\alpha < \epsilon < \frac{\epsilon}{\gamma}$ or $\alpha \in [0, \epsilon) \subset [0, \frac{\epsilon}{\gamma})$. In this case only we can write :

$$\mathrm{E}(M_\alpha^\gamma) < \infty$$

Step 5 : we conclude the proof

The previous steps showed us that, almost surely, the application

$$\begin{aligned}
\omega \colon D &\to \mathbb{R} \\
t &\longmapsto X_t(\omega)
\end{aligned}$$

is uniformly continuous (since it is α Hölder continuous). But since D is dense in $[0, 1]$, we can extend the definition of X by setting :

$$Y_t(\omega) = \begin{cases} X_t(\omega) & t \in D \\ \lim_{\substack{s \to t \\ s \in D}} X_s(\omega) & t \in [0,1] \setminus D \end{cases}$$

so that Y_t is almost surely continuous (we can always impose $Y_t(\omega) = 0$ for all $t \in [0, 1]$ on the null set where Y_t is not continuous).

It remains to show that the stochastic process Y thus obtain is indeed a modification of X, that is :

$$\text{for each fixed } t_0 \in [0,1] \qquad P(\omega \in \Omega : X_{t_0}(\omega) = Y_{t_0}(\omega)) = 1.$$

So we fix $t_0 \in [0,1]$. Let $(t_n) \subset D$ be a sequence such that $\lim t_n = t_0$. Then we have :

$$\begin{aligned} \mathrm{E}[|Y_{t_0} - X_{t_0}|^\gamma] &= \mathrm{E}[\liminf |X_{t_n} - X_{t_0}|^\gamma] \\ &\overset{\text{Fatou}}{\leq} \liminf \mathrm{E}[|X_{t_n} - X_{t_0}|^\gamma] \\ &\overset{\text{hyp}}{\leq} \liminf C \cdot |t_n - t_0|^{1+\epsilon} = 0 \end{aligned}$$

which implies that $|Y_{t_0} - X_{t_0}|^\gamma = 0$ or $Y_{t_0} = X_{t_0}$ almost surely, as desired.

□

Chapter 3

Introduction to Brownian Motion

In this chapter, we construct the pre-Brownian motion following M. Yor and D. Revuz's method (see [1]). We give some of its basic properties and, using the results of the previous chapter, we show the existence of Brownian motion (that is, a pre-Brownian motion with continuous trajectories).

3.1 Motivation

The words "Brownian Motion" correspond to two different ideas : a natural phenomenon and a mathematical object. The natural phenomenon refers to the motion of suspended particles in a liquid. The mathematical object is a Gaussian process for which the variance of an incrementation equals the time elapsed.

It is well known that one of the most efficient ways to model particles at a microscopic level is to consider their movement as random. For a given particle, we can assume that, if its initial position is S_0, its position at time n is $S_n = S_0 + \sum_{i=1}^{n} X_i$ where X_i, $i = 1, 2, ...$ are independent, identically distributed random variables taking values in the space in which the particle is moving (a. e. $\mathbb{R}^d$). Such a process is called a random walk.

It turns out that the characteristics of a microscopic random walk define a macroscopic *picture*, that can, itself, be studied with the tools of probability theory. Initially, Brownian Motion formalism was developed to explain this behaviour, that is the properties of the function $n \mapsto S_n$ from a "global viewpoint" and for a continuous set of indices (representing time). However, this formalism has since been generalised as a model to describe many other phenomena and it is widely used in different areas such as economics, communication theory, management science and -of course- mathematical statistics.

We now give a typical example of a Brownian motion. We are interested in the random movement of a point particle in space. An experiment consists in observing the particle for one second, the outcome of the trial being the (continuous) trajectory of the particle : $\Omega = \mathcal{C}([0,1], \mathbb{R}^3)$. If we denote by Y_t the position of the particle at time t, this experiment can be modelled as :

$$\begin{aligned} Y_t \colon (\Omega, \mathcal{F}, P_X) &\to (\mathbb{R}^3, \mathcal{B}(\mathbb{R}^3)) \\ \omega &\longmapsto Y_t(\omega) = \omega(t) \end{aligned} \tag{3.1}$$

The functions $Y_t, t \in T = [0,1]$ are called the *coordinate mappings*. A point $\omega \in \Omega$ is a continuous application $\omega : [0,1] \longrightarrow \mathbb{R}^3$. The collection containing all the subsets of Ω is obviously a σ-algebra, which we denote by $\mathcal{P}(\Omega)$, so the set of all σ-algebras containing such continuous applications is non empty and we can take the intersection of all its element, which is again a σ-algebra, to obtain the smallest σ-algebra $\mathcal{F}$ for which the continuous applications $t \longmapsto \omega(t)$ are all measurable. Different options are available for defining the probability P_X on $(\Omega, \mathcal{F})$. One of those is the Wiener measure, which will be defined later.

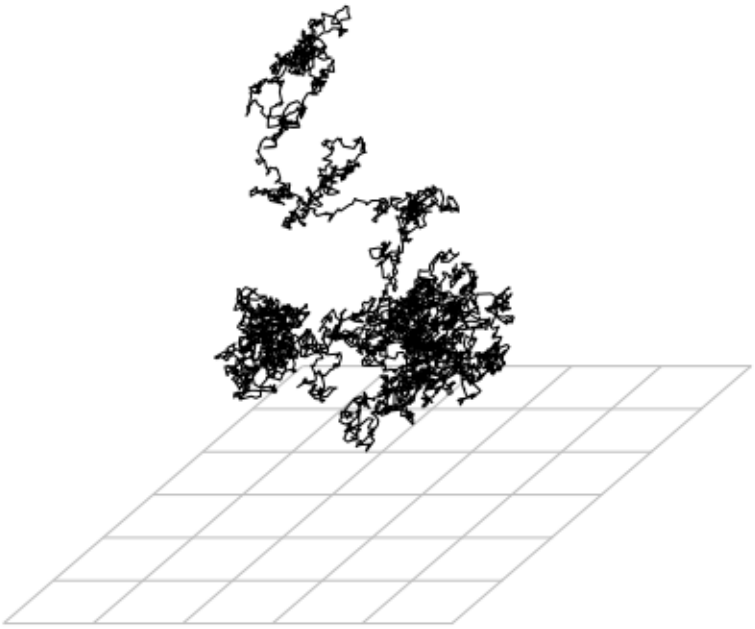

Figure 3.1: A random walk in space

3.2 Construction of a pre Brownian Motion

Formally, a **Brownian motion** (or Wiener process) is a stochastic process $B = (B_t, t \geq 0)$ on a probability space $(\Omega, \mathcal{F}, P)$ that has the following properties :

1. for all finite sequence $t_0 = 0 < t_1 < t_2 < ... < t_n$, the random variables $B_{t_0}, B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, ..., B_{t_n} - B_{t_{n-1}}$ are independent
2. for $0 \leq s, t$, the increment $B_{s+t} - B_s$ is normally distributed with mean 0 and variance t :

$$P(\omega \in \Omega : B_{s+t}(\omega) - B_s(\omega) \in A) = \int_A \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx$$

   for all $A \in \mathcal{B}(\mathbb{R})$
3. with probability 1,

$$\omega : t \longmapsto B_t(w) \text{ is continuous}$$

We also impose that the process starts at 0, but this is just a convention:

$$P(\omega \in \Omega : B_0(\omega) = 0) = 1.$$

Now, if we fix $t \in \mathbb{R}_+$ we get a random variable defined by :

$$B_t \colon (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

$$\omega \longmapsto B_t(\omega)$$

While fixing $\omega \in \Omega$ gives rise to a deterministic function defined by :

$$\omega \colon \mathbb{R}_+ \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

$$t \longmapsto B_t(\omega)$$

This function ω is called a **trajectory**. It represents a possible path of the stochastic process, corresponding to a specific outcome $\omega \in \Omega$. Again, we recall that it is possible to consider ω either like

a function $\omega \in \mathcal{C}([0,1],\mathbb{R}^3)$ or like an outcome $\omega \in \Omega$. The confusion is harmless. Our goal in the following section is to show that such a process exists.

To start the construction of the Brownian motion, we need the following theorem.

Theorem 10. *Given a probability measure μ on $\mathbb{R}$ there exist a probability space $(\Omega,\mathcal{F},P)$ and a sequence of independent random variables $X_n : (\Omega,\mathcal{F},P) \longmapsto (\mathbb{R},\mathcal{B}(\mathbb{R}))$, such that $\mu(A) = P(X_n \in A) = P({X_n}^{-1}(A))$, for all μ-measurable set A and $n \in \mathbb{N}$. This means that μ is the image-measure of P by X_n for all n.*

Proof. See [5], theorem 20.4 p 265. □

To fulfill the construction of the Brownian motion, we first define a Gaussian measure, and show the main properties of such object. Then we define a stochastic process for wich each term (meaning that $t \in \mathbb{R}_+$ is fixed) is the image, by a Gaussian measure, of a certain function in $L^2(E,\mathcal{E},\mu)$. And we finally prove that this process has the desired properties.

Let $(E,\mathcal{E})$ be a measurable space and μ be a σ-finite measure on E. We define a **Gaussian measure** of intensity μ as a linear isometry :

$$G : L^2(E,\mathcal{E},\mu) \longrightarrow \mathcal{G}$$

where $\mathcal{G}$ is a Gaussian space equipped with the inner product (for all $X,Y \in \mathcal{G}$):

$$\langle X,Y\rangle_{\mathcal{G}} = E(XY)$$

while the inner product on the Hilbert space $L^2(E,\mathcal{E},\mu)$ is :

$$\langle f,g\rangle_{L^2(\mu)} = \int_E fg d\mu$$

for all $f,g \in L^2(E,\mathcal{E},\mu)$.

We note that since G is an isometry, we must have in particular $||G(f)||^2_{\mathcal{G}} = ||f||^2_{L^2(\mu)}$ so :

$$||G(f)||^2_{\mathcal{G}} = E(G(f)^2) = \mathrm{Var}(G(f)) = ||f||^2_{L^2(\mu)}. \tag{3.2}$$

Recall that $E(G(f)) = 0$ since $G(f) \in \mathcal{G}$ and $\mathcal{G}$ is a Gaussian space. Using the linearity property of G, we have $\forall f,g \in L^2(E,\mathcal{E},\mu)$:

$$\begin{aligned} E(G(f)^2) - 2E(G(f)G(g)) + E(G(g)^2) &= E((G(f)-G(g))^2) \\ &= E(G(f-g)^2) \\ &= ||f-g||^2_{L^2(\mu)} \\ &= ||f||^2_{L^2(\mu)} - 2\langle f,g\rangle_{L^2(\mu)} + ||g||^2_{L^2(\mu)} \end{aligned}$$

and by simplifying on both sides :

$$\mathrm{Cov}(G(f),G(g)) = E(G(f)G(g)) = \langle f,g\rangle_{L^2(\mu)}. \tag{3.3}$$

For any set $A \in \mathcal{E}$, 1_A is by definition measurable and

$$||1_A||^2_{L^2(\mu)} = \langle 1_A,1_A\rangle_{L^2(\mu)} = \int_E 1_A^2 d\mu = \int_A d\mu = \mu(A)$$

So, if we note $G(A)$ for $G(1_A)$, we have (by 3.2):

$$E(G(A)^2) = ||1_A||^2_{L^2(\mu)} = \mu(A) \Longrightarrow G(A) \sim \mathcal{N}(0,\mu(A)). \tag{3.4}$$

Applying these results to a finite disjoint collection $A_1,A_2,...,A_n \in \mathcal{E}$ of finite μ-measure sets we get (for $i \neq j$):

$$\begin{aligned} E(G(A_i)G(A_j)) &= \langle 1_{A_i},1_{A_j}\rangle_{L^2(\mu)} \\ &= \int_E 1_{A_i}1_{A_j} d\mu \\ &= \int_{A_i\cap A_j} d\mu \\ &= \mu(A_i \cap A_j) = 0 \end{aligned}$$

Since in $G(A_i), G(A_j)$ are gaussian, we have that orthogonality is equivalent to independence and thus :

Proposition 11. Let $A_1, A_2, ..., A_n$ be a finite collection of finite μ-measure sets and G a Gaussian measure of intensity μ. Then :

$$A_1, A_2, ..., A_n \text{ pairwise disjoint } \implies G(A_1), G(A_2), ..., G(A_n) \text{ independent.}$$

We now justify the name Gaussian "measure". Let A be a μ-finite measured set so that :

$$A = \bigsqcup_{i=1}^{\infty} A_i \qquad \text{(disjoint union).}$$

where each A_i is μ-measurable. By defining $\tilde{A}_n = \bigsqcup_{i=1}^{n} A_i$ we observe that $\mu(\tilde{A}_n) = \sum_{i=1}^{n} \mu(A_i) \leq \mu(A) < \infty$ and $G(\tilde{A}_n) = G(1_{\bigsqcup_{i=1}^{n} A_i}) = G(\sum_{i=1}^{n} 1_{A_i}) = \sum_{i=1}^{n} G(A_i)$ for all $n \in \mathbb{N}$ so :

$$\begin{aligned}
\lim_n ||1_{\tilde{A}_n} - 1_A||^2_{L^2(\mu)} &= \lim_n \int_E (1_{\tilde{A}_n} - 1_A)^2 d\mu \\
&= \lim_n \int_E (1_{\bigsqcup_{i=n+1}^{\infty} A_i})^2 d\mu \\
&= \lim_n \int_{\bigsqcup_{i=n+1}^{\infty} A_i} d\mu \\
&= \lim_n \sum_{i=n+1}^{\infty} \mu(A_i) \\
&= 0 \qquad \text{(since the series converges)}
\end{aligned}$$

This implies that :

$$\lim_n ||G(\tilde{A}_n) - G(A)||^2_{\mathcal{G}} = \lim_n ||1_{\tilde{A}_n} - 1_A||^2_{L^2(\mu)} = 0.$$

So we have $G(\tilde{A}_n) = \sum_{i=1}^{n} G(A_i) \longrightarrow G(A)$ for the L^2 convergence. We thus have the following result.

Proposition 12. Let A be a μ-finite measured set so that $A = \bigsqcup_{i=1}^{\infty} A_i$ (disjoint union) where each A_i is μ-measurable. Then :

$$G(A) = G(\bigsqcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} G(A_i)$$

where the convergence holds in the L^2 sense.

We have defined the notion of Gaussian measure. We now need to show how to construct one in an explicit way.

Theorem 13. *Let $H = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ be the separable Hilbert space of square-integrable functions on $\mathbb{R}$, where λ is the Lebesgue measure, which is σ-finite. Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a Gaussian measure $G : L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda) \longrightarrow \mathcal{G}$, where $\mathcal{G}$ is some Gaussian subspace of $L^2(\Omega, \mathcal{F}, P)$.*

Proof. Since H is separable we can pick an orthonormal basis $(e_i, i \in \mathbb{N})$ of H. For all $f \in H$, we have :

$$f = \sum_{i=1}^{\infty} \langle f, e_i \rangle_{L^2(\lambda)} e_i.$$

Recall that the Parseval identity gives us $||f||^2_{L^2(\lambda)} = \sum_{i=1}^{\infty} |\langle f, e_i \rangle_{L^2(\lambda)}|^2 < \infty$. Now, we can use theorem 10 with the probability μ on $\mathbb{R}$ defined by :

$$\mu(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

for all $A \in \mathcal{B}(\mathbb{R})$. The theorem 10 states that there exist a probability space $(\Omega, \mathcal{F}, P)$ and a sequence of independent random variables $X_i : (\Omega, \mathcal{F}, P) \longmapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mu(A) = P(X_i \in A)$ for all $A \in \mathcal{B}(\mathbb{R})$ which is the same as saying that these random variables are reduced normal variables :

$X_i \sim \mathcal{N}(0,1)$ for all $i = 1, 2, ...$ Note that $X_i \sim \mathcal{N}(0,1) \Longrightarrow \langle f, e_i\rangle_{L^2(\lambda)} X_i \sim \mathcal{N}(0, |\langle f, e_i\rangle_{L^2(\lambda)}|^2)$. We can now define what will be our Gaussian measure :

$$G(f) = \sum_{i=1}^{\infty} \langle f, e_i\rangle_{L^2(\lambda)} X_i$$

Since $X_i \in L^2(P)$, $L^2(P)$ is a Hilbert space, $||f||^2_{L^2(\lambda)} = \sum_{i=1}^{\infty} |\langle f, e_i\rangle_{L^2(\lambda)}|^2 < \infty$ and (X_i) is an orthonormal family of $L^2(P)$, we have by a result of functional analysis that the above series converges in $L^2(P)$. If we note $\mathcal{G}$ the Gaussian subspace of $L^2(P)$ generated by (X_i), we have that $\mathcal{G}$ is closed and thus $G(f) \in \mathcal{G}$. Also, by theorem 29, we have that $E(G(f)) = 0$ and $\mathrm{Var}(G(f)) = E(G(f)^2) = \lim_n \sum_{i=1}^{n} |\langle f, e_i\rangle_{L^2(\lambda)}|^2 = ||f||^2_{L^2(\lambda)}$. The linearity of G follows from the properties of the inner product on $L^2(\lambda)$ as well as the isometry property :

$$E(G(f)), ||f||^2_{L^2(\lambda)} \geq 0, E(G(f)^2) = ||f||^2_{L^2(\lambda)} \Longrightarrow ||G(f)||_{\mathcal{G}} = (E(G(f)^2))^{1/2} = ||f||_{L^2(\lambda)}.$$

By linearity of G, we have that $||G(f) - G(g)||_{\mathcal{G}} = ||G(f-g)||_{\mathcal{G}} = ||f-g||_{L^2(\lambda)}$.

So we have proved that $G : L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda) \longrightarrow \mathcal{G}$ is a linear isometry.

□

Remark We can show (using Kolmogorov's three series theorem, for instance) that the above series converges almost surely :

$$P(\omega \in \Omega : G(f)(\omega) = \sum_{i=1}^{\infty} \langle f, e_i\rangle_{L^2(\lambda)} X_i(\omega) < +\infty) = 1.$$

If N_f denotes the set of measure 0 on which the series diverges, we can always define :

$$G(f)(\omega) = 0 \qquad\qquad \forall \omega \in N_f.$$

So without loss of generality, we shall suppose that $G(f)$ converges everywhere. Also, we point out the fact that it would be more accurate to think of $G(f)$ more like an equivalence class than a random variable. Every random variable $G'(f)$ equal to $G(f)$ almost surely would be equivalent to it.

We can now take the first step towards the construction of Brownian motion. Let $H = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ be, as above, the Hilbert space of square-integrable functions on $\mathbb{R}$. If G is the Gaussian measure constructed in the previous theorem, with the assumption that it converges everywhere (see the remark above) then we can define a **pre-Brownian motion**:

$$B_t = G([0,t])$$

where $1_{[0,t]} \in L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ for all $t \geq 0$. Explicitly, we have :

$$B_t = \sum_{i=1}^{\infty} (\int_0^t e_i(x)dx) g_i \quad .$$

Let $t_1 < t_2$ be two positive numbers. Then, assuming that both B_{t_1} and B_{t_2} converge we have:

$$B_{t_2} - B_{t_1} = G([0,t_2]) - G([0,t_1]) = G(]t_1, t_2])$$

Proposition 14. The pre-Brownian motion as we defined it above verifies that for all finite sequence $t_0 = 0 < t_1 < t_2 < ... < t_n$, the random variables $B_{t_0}, B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, ..., B_{t_n} - B_{t_{n-1}}$ are independent.

Proof. Let's define $A_0 = \{0\}$, $A_1 =]0, t_1]$, $A_2 =]t_1, t_2]$, ..., $A_n =]t_{n-1}, t_n]$. We then have that $A_1, A_2, ..., A_n$ is a finite disjoint collection of finite λ-measure sets, so since by construction :

$$B_{t_0} = G(A_0), B_{t_1} - B_{t_0} = G(A_1), B_{t_2} - B_{t_1} = G(A_2), ..., G(A_n) = B_{t_n} - B_{t_{n-1}}$$

we necessarily that the increments are independent (using proposition 11). □

We thus have the first property of a Brownian motion. We are now going to verify the second one. But first, state a very direct consequence of the previous proposition. Since for all finite sequence $t_0 = 0 < t_1 < t_2 < ... < t_n$, the random variables $B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, ..., B_{t_n} - B_{t_{n-1}}$ are independent and normally distributed, we have that their linear transform are also normally distributed. In particular, $(B_{t_1}, B_{t_2}, ..., B_{t_n})$ has the multivariate normal distribution. So **the finite dimensional distributions of the pre-Brownian motion are gaussian**.

Proposition 15. For $0 \leq s, t$, the increment $B_{s+t} - B_s$ is normally distributed with mean 0 and variance t :

$$P(\omega \in \Omega : B_{s+t}(\omega) - B_s(\omega) \in A) = \int_A \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx$$

for all $A \in \mathcal{B}(\mathbb{R})$

Proof. First, we note that by 3.4 we have that :

$$\begin{aligned} \mathrm{Var}(B_{s+t} - B_s) &= \mathrm{Var}(G([0, s+t]) - G([0, s])) \\ &= \mathrm{Var}(G(]s, s+t])) \\ &= E(G(]s, s+t])^2) \\ &= t \end{aligned}$$

and by the linearity of the mean, we must have $E(B_{s+t} - B_s) = 0$. Now since B_{s+t} and B_s are both Gaussian variables, their sum also is (even if they are correlated). □

3.3 Continuity of Paths : Existence of Brownian Motion

As said above, the Brownian motion has to have continuous paths (or trajectories), even though those are very "wiggly". And actually, with our construction of Brownian motion, we can even prove that its trajectories are (almost surely) differentiable nowhere.

To get the continuity, we will use the main result of the previous chapter : the Kolmogorov's continuity criterion.

Theorem 16. *Brownian motion does exist.*

Proof. In the previous section, we were able to construct a stochastic process, the pre-Brownian motion, that has all the properties of the Brownian Motion except continuity. What we want now is to prove that our pre-Brownian motion has a modification that is almost surely hölder-continuous of order α for any $\alpha \in [0, \frac{1}{2})$ (it would be, in particular, continuous). Proposition 15 tells that for all $s, t \in [0, 1]$ with $s < t$, $B_t - B_s \sim |t - s|^{\frac{1}{2}} \cdot \mathcal{N}(0, 1)$ and since all the moments of the law $\mathcal{N}(0, 1)$ are finite, we have that $\mathrm{E}(|N|^\gamma) < \infty$ for all $\infty > \gamma > 2$. We can then define $C_\gamma = \mathrm{E}(|N|^\gamma)$. Also, since $\infty > \gamma > 2$, $\epsilon_\gamma := \frac{\gamma}{2} - 1$ is a well defined strictly positive number and we have $\frac{\gamma}{2} = 1 + \epsilon_\gamma$. Now, we may observe :

$$\begin{aligned} \mathrm{E}(|B_t - B_s|^\gamma) &= \mathrm{E}(|t - s|^{\frac{\gamma}{2}} \cdot |N|^\gamma) \\ &= C_\gamma \cdot |t - s|^{\frac{\gamma}{2}} \\ &= C_\gamma \cdot |t - s|^{1+\epsilon_\gamma} \end{aligned}$$

so the hypothesis of the Kolmogorov continuity criterion are verified. Note that Kolmogorov's criterion asks only for the existence of one such γ. Here, we satisfy the hypothesis for all $\infty > \gamma > 2$. For a fixed γ, we know that there exist a continuous modification that is hölder-continuous of order α for any $\alpha \in [0, \frac{\epsilon_\gamma}{\gamma})$. But since $\frac{\epsilon_\gamma}{\gamma} = \frac{\frac{\gamma}{2} - 1}{\gamma} = \frac{1}{2} - \frac{1}{\gamma}$, we can always change the value of γ to make $\frac{\epsilon_\gamma}{\gamma}$ as close as we want to $\frac{1}{2}$. And thus, we have the α-Hölder continuity for all $\alpha \in [0, \frac{1}{2})$. □

3.4 Wiener Space

We showed (chapter 1) that given a system of finite dimensional distributions (which is a projective family), we could construct a probability space and a process $(X_t, t \in \mathbb{R}_+)$ with the desired finite

dimensional distributions (Kolmogorov extension theorem). We had defined the functions ϕ and Y_t as follows :

$$\begin{aligned} X_t \colon (\Omega, \mathcal{F}, P) &\xrightarrow{\quad\phi\quad} (\mathbb{R}^{\mathbb{R}_+}, \mathcal{R}^{\mathbb{R}_+}) \xrightarrow{\quad Y_t \quad} (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ \omega &\longmapsto (\omega : t \longmapsto X_t(\omega)) \longmapsto \omega(t) = X_t(\omega). \end{aligned}$$

and the law of the process $X = (X_t, t \in \mathbb{R}_+)$ was the image, by ϕ, of the probability P. The application ϕ was measurable because we had equipped $\mathbb{R}^{\mathbb{R}_+}$ with a specific σ-agebra for which the coordinate mappings $Y_t, t \in \mathbb{R}_+$ (for the process X) were measurable. Again, we recall that the confusion arising while considering ω as a point of Ω *and* as a trajectory in $\mathbb{R}^{\mathbb{R}_+}$ is of little harm in practice. So we wilingly do not change the notation.
Now, we wanted to construct a particular process, with given finite dimensional distributions, the Brownian motion. As a consequence of its definition, the finite dimensional distributions of the Brownian motion are normal. Proposition 14 allowed us to show that this was verified for our specific construction of this process. Also, it has (almost surely) continuous trajectories. If $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ represents the continuons functions from $\mathbb{R}_+$ to $\mathbb{R}$, the Brownian motion $B = (B_t, t \in \mathbb{R}_+)$ allows us to redefine the functions ϕ and Y_t by :

$$\begin{aligned} B_t \colon (\Omega, \mathcal{F}, P) &\xrightarrow{\quad\phi\quad} (\mathcal{C}(\mathbb{R}_+, \mathbb{R}), \mathcal{C}^{\mathbb{R}_+}) \xrightarrow{\quad Y_t \quad} (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ \omega &\longmapsto (\omega : t \longmapsto B_t(\omega)) \longmapsto \omega(t) = B_t(\omega). \end{aligned}$$

But what is $\mathcal{C}^{\mathbb{R}_+}$? Well it is again the smallest σ-agebra for which the coordinate mappings $Y_t, t \in \mathbb{R}_+$ (for the process B) are measurable. Continuing the analogy, we have that the law of the Brownian motion $B = (B_t, t \in \mathbb{R}_+)$ is the image, by ϕ, of the probability P. This law, which is a probability measure on $\mathcal{C}^{\mathbb{R}_+}$, is called the **Wiener measure**, and it is denoted by W. And as before, $(Y_t, t \in \mathbb{R}_+)$ is a *canonical* stochastic process. It is a Brownian motion on the space $(\mathcal{C}(\mathbb{R}_+, \mathbb{R}), \mathcal{C}^{\mathbb{R}_+}, W)$ which is called the **Wiener space**.

Chapter 4

Properties of Brownian Motion

4.1 Non Differentiability of Brownian Motion

We mentionned in section 3.3 that it was important to have the *continuity* of paths for a Brownian motion $(B_t, t \in \mathbb{R}_+)$ even if these paths were "wiggly". In this section, we show that they are so wiggly indeed that they are not differentiable. But what do we mean by that ? Does it mean that if we fix $t_0 \in \mathbb{R}_+$, then with probability one the application $t \longmapsto B_t(\omega)$ is not differentiable at time t_0 ? Or do we mean that with probability one, the application $t \longmapsto B_t(\omega)$ is nowhere differentiable ?

Well, both properties are true. And in fact, since the second implies the first, that is the only one we are proving here.

Theorem 17. *Let $(B_t, t \in \mathbb{R}_+)$ be a Brownian motion on Ω. If, as in section 3.4, we think of ω as a specific trajectory $(B_t(\omega), t \in \mathbb{R}_+)$ of the considered Brownian motion, then the subset D of Ω containing the (somewhere) differentiable trajectories is contained in a set of measure 0.*

Proof. The proof contains three main steps.

Step 1 : review of a simple result about differentiable functions

Let $g : \mathbb{R}_+ \longmapsto \mathbb{R}$ be a differentiable function at point $t \in \mathbb{R}_+$, (here, we fix such t). Then if (h_n) is any sequence such that $\lim h_n = 0$, we have :

$$\lim_{n\to\infty} \frac{|g(t+h_n) - g(t)|}{|h_n|} < +\infty \tag{4.1}$$

In particular, if (h_n) is a positive sequence, we can state that there exists $l \in \mathbb{N}$ and $N \in \mathbb{N}$ such that :

$$|g(t+h_n) - g(t)| \leq l \cdot h_n \qquad \text{for all } n > N. \tag{4.2}$$

Now, let $i = [nt] + 1$, then $i \in \{1, 2, ..., n, n+1\}$ and we have :

$$\begin{aligned} [nt] \leq \quad & nt \quad && \leq [nt] + 1 \\ \frac{[nt]}{n} \leq \quad & t \quad && \leq \frac{[nt]}{n} + \frac{1}{n} \\ \frac{[nt]+1}{n} - \frac{1}{n} \leq \quad & t \quad && \leq \frac{[nt]+1}{n} \\ \frac{i}{n} - \frac{1}{n} \leq \quad & t \quad && \leq \frac{i}{n} \\ 0 \leq \quad & \tfrac{i}{n} - t \quad && \leq \frac{1}{n} \end{aligned} \tag{4.3}$$

which ensures that $\lim_{n\to\infty} \frac{i}{n} = t$.
Secondly, for a fixed i, let j be any number in $\{i+1, i+2, i+3\}$. Then, $j - i \in \{1, 2, 3\}$ so we can add $\frac{j-i}{n} \geq 0$ to the last inequality above to get :

$$\begin{aligned} 0 \leq \quad & \tfrac{i}{n} - t \quad && \leq \frac{1}{n} \\ 0 \leq \quad & \tfrac{j-i}{n} + \tfrac{i}{n} - t \quad && \leq \frac{j-i}{n} + \frac{1}{n} \\ 0 \leq \quad & \tfrac{j}{n} - t \quad && \leq \frac{j-i+1}{n} \leq \frac{4}{n} \end{aligned} \tag{4.4}$$

So we have :

$$0 \leq \frac{j}{n} - t \leq \frac{4}{n} \qquad \text{and} \qquad 0 \leq \frac{j-1}{n} - t \leq \frac{3}{n}$$

which ensures that $\lim_{n\to\infty} \frac{j}{n} = \lim_{n\to\infty} \frac{j-1}{n} = t$.

All of these results were proved to verify :

$$\begin{aligned} |g(\frac{j}{n}) - g(\frac{j-1}{n})| \quad & \leq |g(\tfrac{j}{n}) - g(t)| + |g(\tfrac{j-1}{n}) - g(t)| \\ & \leq l \cdot (\tfrac{j}{n} - t) + l \cdot (\tfrac{j-1}{n} - t) \quad \leq \frac{4l}{n} + \frac{3l}{n} = \frac{7l}{n} = \frac{l'}{n} \qquad \text{(for } n \text{ sufficiently large)} \end{aligned} \tag{4.5}$$

where we used 4.2 and noted $l' = 7l \in \mathbb{N}$.

Step 2 : definition of Γ and proof that $D \subset \Gamma$

First we define :

$$D_{t_0} = \{\omega \in \Omega : t \longmapsto B_t(\omega) \text{ differentiable at } t_0 \in \mathbb{R}_+\} \tag{4.6}$$

With this definition, it is clear that $D = \{\omega \in \Omega : \exists t \in \mathbb{R}_+ \text{ verifying } t \longmapsto B_t(\omega) \text{ differentiable at } t\} = \bigcup_t D_t$. Now, let's fix $\omega \in D$, there must exist, by definition, a $t \in \mathbb{R}_+$ such that $t \longmapsto B_t(\omega)$ is differentiable at t. For short, we shall write $B(t)$ instead of $B_t(\omega)$, since ω is fixed. Using step 1, we can write :

$$\exists l \in \mathbb{N}, \exists N \in \mathbb{N} : \forall n > N, \exists i \in \{1, 2, .., n, n+1\}, \forall j \in \{i+1, i+2, i+3\} : |B(\frac{j}{n}) - B(\frac{j-1}{n})| \leq \frac{l}{n}.$$

But this is exactly the same as writing (since we took any $\omega \in D = \bigcup_t D_t$) :

$$\bigcup_t D_t \subset \Gamma = \bigcup_{l\in\mathbb{N}} \bigcup_{N\in\mathbb{N}} \bigcap_{n>N} \bigcup_{i=1}^{n+1} \bigcap_{i+1}^{i+3} \{\omega \in \Omega : |B(\frac{j}{n}) - B(\frac{j-1}{n})| \leq \frac{l}{n}\} \tag{4.7}$$

Step 3 : proof that $P(\Gamma) = 0$

First, we note that for all fixed $i \in \{1, 2, ..., n, n+1\}$, we can define

$$t_1 = \frac{i}{n} \qquad t_2 = \frac{i+1}{n} \qquad t_3 = \frac{i+2}{n} \qquad t_4 = \frac{i+3}{n}$$

so that $t_k - t_{k-1} = \frac{1}{n}$ for all $k \in \{2, 3, 4\}$. Then :

$$\begin{aligned} P(\textstyle\bigcap_{j=i+1}^{i+3} \{|B(\frac{j}{n}) - B(\frac{j-1}{n})| \leq \frac{l}{n}\}) \quad & = P(\bigcap_{k=2}^{4} \sqrt{n}|B_{t_k} - B_{t_{k-1}}| \leq \frac{l}{\sqrt{n}}) \\ & \overset{Prop.14}{=} \prod_{k=2}^{4} P(\sqrt{n}|B_{t_k} - B_{t_{k-1}}| \leq \frac{l}{\sqrt{n}}) \\ & \overset{Prop.15}{=} \prod_{k=2}^{4} P(|\mathcal{N}(0,1)| \leq \frac{l}{\sqrt{n}}) \leq \prod_{k=2}^{4} \frac{l}{\sqrt{n}} = \frac{l^3}{n^{\frac{3}{2}}} \end{aligned} \tag{4.8}$$

where the last inequality follows from the normal distribution properties. Thus, we can write :

$$\begin{aligned} P(\bigcap_{n=N+1}^{\infty}\bigcup_{i=1}^{n+1}\bigcap_{j=i+1}^{i+3}\{|B(\tfrac{j}{n})-B(\tfrac{j-1}{n})|\leq\tfrac{l}{n}\}) \\ = P(\bigcap_{n=N+1}^{\infty}A_n) \\ \leq \inf_{n\geq N}P(A_n) \\ = \inf_{n\geq N}P(\bigcup_{i=1}^{n+1}\bigcap_{j=i+1}^{i+3}\{|B(\tfrac{j}{n})-B(\tfrac{j-1}{n})|\leq\tfrac{l}{n}\}) \\ \leq \inf_{n\geq N}\sum_{i=1}^{n+1}P(\bigcap_{j=i+1}^{i+3}\{|B(\tfrac{j}{n})-B(\tfrac{j-1}{n})|\leq\tfrac{l}{n}\}) \\ \leq \inf_{n\geq N}\tfrac{n+1}{n^{\frac{3}{2}}}l^3=0 \end{aligned} \tag{4.9}$$

This results says that Γ is the countable union of sets of probability 0. Then, the set Γ itself must be of probability 0 : $P(\Gamma)=0$, which ends the proof since $D\subset\Gamma$.

□

4.2 Weak Markov Property

There exist two versions of the Markov property, a weak and a strong version. Now, obviously the strong Markov property is stronger that the weak Markov property but unfortunately, we cannot proof the former without the latter. So we first proof the weak Markov property.

Theorem 18 (Weak Markov Property)**.** *Let $(B_t, t\in\mathbb{R}_+)$ be a Brownian motion. And fix $t_0\in\mathbb{R}_+$. Then we have that :*

$$B'_t = B_{t_0+t}-B_{t_0} \qquad\qquad t\geq 0$$

is a Brownian motion independent of $\sigma(B_s, s\leq t_0)$.

Proof. Consider the sequences

$$\begin{aligned} 0\leq s_1\leq s_2\leq ...\leq s_j \qquad & \text{with } s_l\leq t_0 \quad 1\leq l\leq j \\ 0\leq t_1\leq t_2\leq ...\leq t_k \qquad & \text{with no restriction on the } t_l \end{aligned}$$

Then we can set $u_i=t_0+t_i$ (note that we have $0\leq s_1\leq s_2\leq ...\leq s_j\leq u_1\leq u_2\leq ...\leq u_k$). Since the increments of Brownian motion are independant (See prop. 14), we have :

$$(B_{u_1}-B_{t_0}, B_{u_2}-B_{t_1}, ..., B_{u_k}-B_{t_{k-1}}) \quad \text{independent of} \quad (B_{s_1}, B_{s_2}-B_{t_1}, ..., B_{s_j}-B_{s_{j-1}}) \tag{4.10}$$

But $B_{u_1}-B_{t_0}=B_{t_0+t_1}-B_{t_0}=B'_{t_1}$ and more generally

$$B_{u_l}-B_{u_{l-1}}=B_{t_0+t_l}-B_{t_0+t_{l-1}}=(B_{t_0+t_l}-B_{t_0})-(B_{t_0+t_{l-1}}-B_{t_0})=B'_{t_l}-B'_{t_{l-1}}$$

for $2\leq l\leq k$. So 4.11 gives

$$(B'_{t_1}, B'_{t_2}-B'_{t_1}, ..., B'_{t_k}-B'_{t_{k-1}}) \quad \text{independent of} \quad (B_{s_1}, B_{s_2}-B_{t_1}, ..., B_{s_j}-B_{s_{j-1}}) \tag{4.11}$$

And, using the $\pi-\lambda$ theorem (see theorem 2), we have the independence of $(B'_{t_1}, B'_{t_2}-B'_{t_1}, ..., B'_{t_k}-B'_{t_{k-1}})$ and $\sigma(B_s, s\leq t_0)$. Also, we have :

1. $B'_0=B_{t_0+0}-B_{t_0}=0$
2. $B'_t=B_{t_0+t}-B_{t_0}\sim\mathcal{N}(0,t)\Longrightarrow B'_{t_2}-B_{t_1}=B_{t_0+t_2}-B_{t_0+t_1}\sim\mathcal{N}(0,t_2-t_1)$
3. The increments are independent

So $(B'_t, t\in\mathbb{R}_+)$ is a Brownian motion independent of $\sigma(B_s, s\leq t_0)$. As a consequence we have that for any $H\in\mathcal{B}(\mathbb{R}^k)$ and $A\in\sigma(B_s, s\leq t_0)$, we have :

$$\begin{aligned} P(((B'_{t_1}, B'_{t_2}, ..., B'_{t_k})\in H)\cap A) \quad &= P((B'_{t_1}, B'_{t_2}, ..., B'_{t_k})\in H)P(A) \\ &= P((B_{t_1}, B_{t_2}, ..., B_{t_k})\in H)P(A) \end{aligned} \tag{4.12}$$

where the last equality comes from the fact that $(B'_t, t\in\mathbb{R}_+)$ has the same distribution as $(B_t, t\in\mathbb{R}_+)$, that is, the distribution of the Brownian motion. □

4.3 Strong Markov Property

Having established the result of the previous section, we have the necessary tools for proving the strong Markov property.

Theorem 19 (Strong Markov Property)**.** *Let $(B_t, t \in \mathbb{R}_+)$ be a Brownian motion and τ a stopping time. Then we have that :*

$$B_t^* = B_{\tau+t} - B_\tau \qquad\qquad t \geq 0$$

is a Brownian motion independent of $\mathcal{F}_\tau = \{M \text{ measurable} : M \cap \{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t \ \forall t \geq 0\}$ with $\mathcal{F}_t = \sigma(B_s, s \leq t)$. In particular,

$$\begin{aligned} P(((B_{t_1}^*, B_{t_2}^*, ..., B_{t_k}^*) \in H) \cap M) &= P((B_{t_1}^*, B_{t_2}^*, ..., B_{t_k}^*) \in H)P(M) \\ &= P((B_{t_1}, B_{t_2}, ..., B_{t_k}) \in H)P(M) \end{aligned} \tag{4.13}$$

for all sequences $t_1 \leq t_2 \leq ... \leq t_k$, $M \in \mathcal{F}_\tau$ and $H \in \mathcal{B}(\mathbb{R}^k)$.

Proof. The proof contains three steps. First of all, we consider a special case.

Step 1 : τ has a countable image

In the case where τ has a countable image V, we have :

$$\{\omega \in \Omega : B_t^*(\omega) \in H\} = \textstyle\bigcup_{t_0 \in V} \{\omega \in \Omega : B_{t_0+t} - B_{t_0} \in H, \tau(\omega) = t_0\} \tag{4.14}$$

Since, τ is a stopping time, we have by definition $\{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. And since $\{\omega \in \Omega : \tau(\omega) = t\} = \{\omega \in \Omega : \tau(\omega) \leq t\} \setminus \bigcup_{n \in \mathbb{N}} \{\omega \in \Omega : \tau(\omega) \leq t - \frac{1}{n}\}$ for all $t \geq 0$, we have that $\{\omega \in \Omega : \tau(\omega) = t\} \in \mathcal{F}_t$. Also, it is well known that the intersection of two measurable sets is measurable. Thus, the union above is the union of measurable sets and it is itself measurable. We have : $\{\omega \in \Omega : B_t^*(\omega) \in H\} \in \mathcal{F} = \sigma(B_t, t \in \mathbb{R}_+)$. This means in particular that B_t^* is a random variable.

Now, let $M \in \mathcal{F}_\tau$ and $H \in \mathcal{B}(\mathbb{R}^k)$. We have :

$$\begin{aligned} P(((B_{t_1}^*, B_{t_2}^*, ..., B_{t_k}^*) \in H) \cap M) &= \textstyle\sum_{t_0 \in V} P(((B_{t_1}^*, B_{t_2}^*, ..., B_{t_k}^*) \in H) \cap M \cap (t = t_0)) \\ &= \textstyle\sum_{t_0 \in V} P(((B_{t_1}^{'}, B_{t_2}^{'}, ..., B_{t_k}^{'}) \in H) \cap M \cap (t = t_0)) \\ &\overset{Th.18}{=} P((B_{t_1}^{'}, B_{t_2}^{'}, ..., B_{t_k}^{'}) \in H) \textstyle\sum_{t_0 \in V} P(M \cap (t = t_0)) \\ &= P((B_{t_1}^{'}, B_{t_2}^{'}, ..., B_{t_k}^{'})P(M) \end{aligned} \tag{4.15}$$

Letting $M = \Omega$, we get

$$\begin{aligned} P(((B_{t_1}^*, B_{t_2}^*, ..., B_{t_k}^*) \in H) &= P(((B_{t_1}^*, B_{t_2}^*, ..., B_{t_k}^*) \in H) \cap M) \\ &= P((B_{t_1}^{'}, B_{t_2}^{'}, ..., B_{t_k}^{'})P(M) \\ &= P((B_{t_1}^{'}, B_{t_2}^{'}, ..., B_{t_k}^{'}) \in H) \end{aligned}$$

Injecting this result in 4.15 we have :

$$P(((B_{t_1}^*, B_{t_2}^*, ..., B_{t_k}^*) \in H) \cap M) = P((B_{t_1}^*, B_{t_2}^*, ..., B_{t_k}^*) \in H)P(M)$$

meaning that $\sigma(B_t^*, t \in \mathbb{R}_+)$ and $\mathcal{F}_\tau$ are independent.

Step 2 : general case and approximation of τ

The main idea here is to use a density argument. So we construct a sequence of stopping time (τ_n), which will converge to τ, as follows. Let's fix $n \in \mathbb{N}$. We now define $\tau_n(\omega)$ where $\omega \in \Omega$ is fixed. First of all, $\tau(\omega) \in \mathbb{R}_+$. Since $\mathbb{R}_+^* = (\frac{0}{2^n}, \frac{1}{2^n}] \cup (\frac{1}{2^n}, \frac{2}{2^n}] \cup ... \cup (\frac{k-1}{2^n}, \frac{k}{2^n}] \cup ...$, there exits a *unique* $k \in \mathbb{N}$ so that $\frac{k-1}{2^n} < \tau(\omega) \leq \frac{k}{2^n}$ and we can define $\tau_n(\omega) = \frac{k}{2^n}$. In the case where $\tau(\omega) = 0$, we set $\tau_n(\omega) = 0$. This entirely defines the sequence (τ_n). Note that for any fixed $\omega \in \Omega$, $\tau_n(\omega) \searrow \tau(\omega)$. In particular, (τ_n) is a decreasing sequence.
We now fix $n \in \mathbb{N}$. Our goal is to show that t_n is a stopping time. So we fix $t \in \mathbb{R}_+$, we want to show

that $\{\omega \in \Omega : \tau_n(\omega) \leq t\} \in \mathcal{F}_t$ Again, in a slightly different shape, we have $\mathbb{R}_+ = [\frac{0}{2^n}, \frac{1}{2^n}) \cup [\frac{1}{2^n}, \frac{2}{2^n}) \cup ... \cup [\frac{k-1}{2^n}, \frac{k}{2^n}) \cup ...$ and there exists a unique $k \in \mathbb{N}$ so that $\frac{k}{2^n} \leq t < \frac{k+1}{2^n}$. This allows us to write :

$$\{\omega \in \Omega : \tau_n(\omega) \leq t\} = \{\omega \in \Omega : \tau(\omega) \leq \frac{k}{2^n}\} \in \mathcal{F}_{\frac{k}{2^n}} \subset \mathcal{F}_t$$

so the sequence is a sequence of decreasing stopping times. Since we want to show that $\sigma(B_t, t \in \mathbb{R}_+)$ is independent of $\mathcal{F}_\tau$, we pick $M \in \mathcal{F}_\tau$. We have :

$$M \cap (\tau_n \leq t) = N \cap (\tau \leq \frac{k}{2^n}) \in \mathcal{F}_{\frac{k}{2^n}} \subset \mathcal{F}_t$$

that is to say $M \in \mathcal{F}_\tau \Longrightarrow M \in \mathcal{F}_{\tau_n}$ or $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$. We know define the approximation

$$B_t^n = B_{\tau_n + t} - B_{\tau_n} \tag{4.16}$$

which corresponds to B_t^* for $\tau = \tau_n$. But notice the very important fact that in this case, the image of τ_n is countable. So, using step 1, we have :

$$P(((B_{t_1}^n, B_{t_2}^n, ..., B_{t_k}^n) \in H) \cap M) = P((B_{t_1}^n, B_{t_2}^n, ..., B_{t_k}^n) \in H)P(M) \tag{4.17}$$

for all $M \in \mathcal{F}_{\tau_n}$ and $H \in \mathcal{B}(\mathbb{R}^k)$.

Step 3 : conclusion

We start by recalling the two following facts :

1. $\forall \omega \in \Omega : \lim_{n\to\infty} \tau_n(\omega) = \tau(\omega)$
2. $\forall \omega \in \Omega : t \longmapsto B_t(\omega)$ is continuous.

Also, we quote the following elementary lemma from [5], p502.

Lemma 20. (Billingsley) If the following conditions are satisfied:

1. X_n, X be two k-dimensional vectors
2. $F_n(x) = P(X_n \leq x)$ where $x \in \mathbb{R}^k$
3. $X_n \longrightarrow X$ with probability 1
4. $\forall x \in \mathbb{R}^k$, $\lim_{n\to\infty} F_n(x) = F(x)$

then $F(x) = P(X \leq x)$

To use this lemma here, we set :

- $X_n = (B_{\tau_n+t_1} - B_{\tau_n}, B_{\tau_n+t_2} - B_{\tau_n}, ..., B_{\tau_n+t_k} - B_{\tau_n}) = (B_{t_1}^n, B_{t_2}^n, ..., B_{t_k}^n)$
- $X = (B_{\tau+t_1} - B_\tau, B_{\tau+t_2} - B_\tau, ..., B_{\tau+t_k} - B_\tau) = (B_{t_1}^*, B_{t_2}^*, ..., B_{t_k}^*)$
- $F_n(x) = F_n(x_1, x_2, ..., x_k) = P(B_{t_1}^n \leq x_1, B_{t_2}^n \leq x_2, ..., B_{t_k}^n \leq x_k)$
- $P(\omega \in \Omega : \lim_{n\to\infty}(B_{t_1}^n(\omega), B_{t_2}^n(\omega), ..., B_{t_k}^n(\omega)) = (B_{t_1}^*(\omega), B_{t_2}^*(\omega), ..., B_{t_k}^*(\omega))) = 1$ by continuity.
- Since for all $n \in \mathbb{N}$, (B_t^n) has the distribution of Brownian motion (see step 1), we know the (Gaussian) distribution of $(B_{t_1}^n, B_{t_2}^n, ..., B_{t_k}^n)$ whatever n is. And this distribution is the same for all n because the finite dimensional distributions of Brownian motion are uniquely defined. As a matter of form however, we denote by F_n the (Brownian) distribution of the preceding vector. As we said, $F_n = F$ does not depend on n. So by construction $\lim F_n = F$.

We can now apply the lemma to get :

$$F(x) = P(B_{t_1}^* \leq x_1, B_{t_2}^* \leq x_2, ..., B_{t_k}^* \leq x_k)$$

which says that $(B_{t_1}^*, B_{t_2}^*, ..., B_{t_k}^*)$ has the finite dimensional distribution of Brownian motion. Recall that the equation 4.17 read :

$$P(((B^n_{t_1}, B^n_{t_2}, ..., B^n_{t_k}) \in H) \cap M) = P((B^n_{t_1}, B^n_{t_2}, ..., B^n_{t_k}) \in H)P(M)$$

or

$$P((B_{t_1}, B_{t_2}, ..., B_{t_k}) \in H|M) = P((B_{t_1}, B_{t_2}, ..., B_{t_k}) \in H)$$

since the result of the theorem holds for $(B^n_{t_1}, B^n_{t_2}, ..., B^n_{t_k})$, by step 1. Now, we showed that $(B^*_{t_1}, B^*_{t_2}, ..., B^*_{t_k})$ had the same distribution as the k-dimensional (Brownian) vector, so we have :

$$P(((B^*_{t_1}, B^*_{t_2}, ..., B^*_{t_k}) \in H)|M) = P((B^*_{t_1}, B^*_{t_2}, ..., B^*_{t_k}) \in H) \tag{4.18}$$

and this ends the proof. □

Appendix A

Preliminaries

A.1 Mesurable Functions and Random Variables

The goal of probability theory is to analyse phenomenons giving rise to multiple potential outcomes that, presumably, cannot be *exactly* predicted, but for which **the set of all the possible outcomes**, denoted by Ω, is well defined. Each possible outcome $\omega \in \Omega$ is called an **elementary event**. For a given couple $(\Omega, \mathcal{F})$, we define an **event** as an element of the set $\mathcal{F} \subset \mathcal{P}(\Omega)$, which is a σ-algebra.

If $\mathcal{A}$ is any subset of $\mathcal{P}(\Omega)$, then we define the **σ-algebra generated by $\mathcal{A}$** by taking the intersection of all the σ-algebras containing $\mathcal{A}$ and we denote it by $\sigma(\mathcal{A})$. The family of such σ-algebras is not empty since $\mathcal{P}(\Omega)$ belongs to it. It is easy to verify that an intersection of σ-algebras is again a σ-algebra.

Example 1. The σ-algebra generated by the open sets of a topological space Ω is fundamental for the theory. Its sets are called **Borel sets** and are denoted by $\mathcal{B}(\Omega)$. It is the smallest σ-algebra containing all the open sets of Ω. While dealing with topological spaces, we will assume that they are equipped with this particular σ-algebra.

A couple of sets $(\Omega, \mathcal{F})$ is called a measurable space, this is a space on which we can define a measure. The elements of $\mathcal{F}$ are called **measurable sets**. Now, a **positive measure** is a nonnegative countably additive set function, that is, a function $\mu : \mathcal{F} \longrightarrow [0, \infty]$ with :

1. $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{F}$
2. if $A_i \in \mathcal{F}$ is a countable sequence of *disjoint* sets, then :

$$\mu(\bigcup A_i) = \sum \mu(A_i)$$

A measure μ is said to be **σ-finite** if there exists an increasing and countable sequence of measurable sets $A_1 \subset A_2 \subset A_3 ...$ so that $\Omega = \bigcup A_i$ and $\mu(A_i) < \infty$ for all $i = 1, 2, ...$. If $\mu(\Omega) = 1$, μ is called a **probability measure**. Such measures will usually be denoted by P, to emphasize the fact that, when evaluated at a certain event A, they represent the probability of such event, i. e. the frequency of occurrence of A if we repeat the experiment associated with Ω an infinite number of times in the same initial conditions. Note that the convergence of the proportion of times that A occurs is not proven but is just a belief, confirmed by scientific experimentation. A triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space and, if μ is a probability measure, a **probability space**.

Theorem 21 (Measures on the real line)**.** *To any given function $F : \mathbb{R} \longrightarrow \mathbb{R}$ that is nondecreasing and right continuous, we can associate a measure μ on $(\mathbb{R}, \mathcal{F})$, where $\mathcal{F} = \mathcal{B}(\mathbb{R})$, by :*

$$\mu((a, b]) = F(b) - F(a)$$

for every bounded interval $(a, b]$.

F is called a **Stieltjes measure function**. We note that the σ-algebra generated by the open sets, and the σ-algebra generated by the half opened intervals (of the type $(a, b]$) are the same. Thus, the function F entirely determines the measure μ over $\mathcal{F}$. If we set $F(x) = x$ we obtain the very important **Lebesgue measure**, denoted by λ. It is the only measure defined on $\mathcal{B}(\mathbb{R})$ which associates to any real interval its length.

Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ be two measurable spaces. A function $f : \Omega \longrightarrow E$ is said to be a **measurable function** if for any set $A \in \mathcal{E}$, $f^{-1}(A) \in \mathcal{F}$.

Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ be a probability space and a measurable space respectively. A measurable function $X : \Omega \longrightarrow E$ is called a **random variable**. For now on, unless stated otherwise, we shall assume that $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}(\mathbb{R})$.

Theorem 22. *If $(f_n)_{n\in\mathbb{N}}$ is a sequence of extended real valued measurable functions on a measurable space $(\Omega, \mathcal{F})$, then each of the four following functions is measurable (we fix $\omega \in \Omega$) :*

$$\begin{aligned} h(\omega) &= \sup\{f_n(\omega) : n = 1, 2, ..\}, \\ g(\omega) &= \inf\{f_n(\omega) : n = 1, 2, ..\}, \\ f^*(\omega) &= \limsup_n\{f_k(\omega) : k = n, n + 1, ..\}, \\ f_*(\omega) &= \liminf_n\{f_k(\omega) : k = n, n + 1, ..\}. \end{aligned}$$

More generally the set $\{\omega \in \Omega \mid lim\ f_n(\omega)\ exists\}$ is measurable.

Proof. Let $h(\omega) = \sup\{f_n(\omega) : n = 1, 2, ..\} = \sup f_n(\omega)$. We want to show that for any $a \in \mathbb{R}$, $h^{-1}((-\infty, a)) \in \mathcal{F}$. To do so, we note that :

$$h^{-1}((-\infty, a)) = \{\omega : \sup f_n(\omega) < a\} = \bigcap_n \{\omega : f_n(\omega) < a\}.$$

Using the fact that, for all $n = 1, 2, ...$, f_n is measurable, we have that $\{\omega : f_n(\omega) < a\}$ is measurable and so is $h^{-1}((-\infty, a))$. For g, we use the same reasoning but with $g^{-1}((-\infty, a)) = \bigcup_n \{\omega : f_n(\omega) < a\}$. To conclude, we note that $x_n = \sup_n\{f_k(\omega) : k = n, n + 1, ..\}$ is a decreasing sequence, so $\lim\ x_n = \inf_{n\geq 0}\ x_n$, and we write $f^* = \inf_{n\geq 0}(\sup_{k\geq n} f_k)$, that is, f^* is a composition of measurable functions, so it is measurable. The same holds for f_* but in this case we have $f_* = \sup_{n\geq 0}(\inf_{k\geq n} f_k)$.

□

The previous theorem tells us that

$$\Omega_0 = \{\omega \in \Omega \mid \lim X_n(\omega) \text{ exists}\} = \{\omega \in \Omega \mid \limsup X_n(\omega) = \liminf X_n(\omega)\}$$

is measurable. If its measure is 1, we say that X_n **converges almost surely** (see next section).

If $f : (\Omega, \mathcal{F}) \longrightarrow (E, \mathcal{E})$ is a measurable function and $\mu : (\Omega, \mathcal{F}) \longrightarrow [0, \infty]$ a positive measure on $(\Omega, \mathcal{F})$, then the **image-measure** of μ with respect to f, denoted by $f(\mu)$ is the positive measure on $(E, \mathcal{E})$ defined, for all $A \in \mathcal{E}$, by :

$$f(\mu)(A) = \mu(f^{-1}(A)).$$

As we will see below, given a probability space $(\Omega, \mathcal{F}, P)$ and a random variable X taking its values in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, it is easy to construct a probability measure on $\mathcal{B}(\mathbb{R})$ by taking the image of P by X, namely $X(P)$.

The **distribution** μ_X of a random variable (or random vector) X on a space $(\Omega, \mathcal{F}, P)$ is the image-measure of P with respect to X. It is thus the probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (or $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$) defined $\forall B \in \mathcal{B}(\mathbb{R})$ (or $\mathcal{B}(\mathbb{R}^n)$) by :

$$\mu_X(B) = X(P)(B) = P(X^{-1}(B)) = P(X \in B).$$

Note that for any given probability measure P on the space $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the random variable defined by $X(\omega) = \omega$ has the $P = \mu_X$ distribution. Generalizing this fact, we can consider

$(\Omega, \mathcal{F}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ and let P be a probability measure defined on $\mathcal{B}(\mathbb{R}^n)$. We now define a random vector X on Ω by $X_i(\omega) = \omega_i$. The measure P is then the distribution of $X = (X_1, X_2, ..., X_n)$ since

$$\mu_X(A_1 \times A_2 \times ... \times A_n) = P(X \in A_1 \times A_2 \times ... \times A_n) = P(A_1 \times A_2 \times ... \times A_n)$$

for all $A_1, A_2, ..., A_n \in \mathcal{B}(\mathbb{R})$. But we can do better. Suppose $\mu_1, ..., \mu_n$ is a *finite* sequence of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. As before, we let $X_i(\omega) = \omega_i$ and we define P on $(\Omega, \mathcal{F}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ by :

$$P(A_1 \times A_2 \times ... \times A_n) = \mu_1(A_1)\mu_2(A_2)...\mu_n(A_n)$$

for all $A_1, A_2, ..., A_n \in \mathcal{B}(\mathbb{R})$. Then we have : $\mu_{X_i}(A_i) = P(X_i \in A_i) = P(\omega \in \Omega : \omega_i \in A_i$, and $\omega_j \in \mathbb{R}$ for all $i \neq j) = P(\mathbb{R} \times ... \times A_i \times ... \times \mathbb{R}) = \mu_1(\mathbb{R})...\mu_i(A_i)...\mu_n(\mathbb{R}) = \mu_i(A_i)$ so X_i has the distribution μ_i. And thus :

$$P(A_1 \times A_2 \times ... \times A_n) = P(X \in A_1 \times A_2 \times ... \times A_n) = \mu_1(A_1)\mu_2(A_2)...\mu_n(A_n) = P(X_1 \in A_1)P(X_2 \in A_2)...P(X_n \in A_n)$$

which means that $X_1, X_2, ..., X_n$ are *independent* (see below, section 1.3). To sum up, given a finite sequence of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we were able to construct a probability space $(\Omega, \mathcal{F}, P)$ and an independent sequence of random variables having the given measures as distributions. In this paper, we will try to apply this method to an *infinite* sequence of probability measures.

If two random variables X and Y induce the same distribution μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we say that X and X are **equal in distribution** and we write :

$$X \overset{D}{=} Y.$$

The **distribution function** F_X of a real-valued random variable X is the function defined by :

$$F(x) = P(X \leq x) \qquad \text{for all } x \in \mathbb{R}.$$

The σ-algebra **generated by** some set of functions $(Y_t, t \in T)$ taking their values in a measurable set $(E, \mathcal{E})$ is defined by :

$$\sigma(Y_t, t \in T) = \sigma(Y_t^{-1}(B) \mid B \in \mathcal{E}, t \in T).$$

So, for a given space Ω and some functions defined on it, we can always define a σ-algebra on Ω for which these functions are measurable. This will play an important role when we construct a σ-algebra on the product space $\Omega = \mathbb{R}^{\mathbb{R}_+}$, the set of all functions $\omega : t \longmapsto \omega(t)$ from $\mathbb{R}_+$ to $\mathbb{R}$. In this case, we shall put, for all $t \in T = \mathbb{R}_+$, $Y_t : \mathbb{R}^{\mathbb{R}_+} \longrightarrow \mathbb{R}$, $\omega \longmapsto \omega(t)$. The functions $(Y_t, t \in \mathbb{R}_+)$ will be called *coordinate mappings*.

Suppose that f is a nonnegative measurable function on a measure space $(E, \mathcal{E})$ equipped with the measure μ. We can define a measure ν by

$$\nu(A) = \int_A f d\mu$$

for all measurable subset A of E. The properties of the integral ensure that ν is indeed a measure. In particular, we see that $\mu(A) = 0 \Longrightarrow \nu(A) = 0$ and ν is finite if and only if f is μ-integrable. Also, it is clear that if $f = f'$ a.e. then $\nu = \nu'$. With such definition, the measure ν is said to have **density** f with respect to μ. A density is by definition nonnegative and if it is μ-integrable, we can always normalise it to guarantee $\nu(E) = 1$, and make ν become a probability measure on E (which will be thought of as the distribution of some random variable X). In the case where $E = \mathbb{R}$ and $\mu = \lambda$, we have the well known situation where f is the Lebesgue-density of the distribution ν.
We give here a very famous example of such construction, where we define a random variable through its μ-density. Consider the sets $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $(E, \mathcal{E}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. We let $q : x \longmapsto \exp(-\frac{x^2}{2})$ be a nonnegative measurable function on $E = \mathbb{R}$. Since q is λ-integrable and $\int q d\lambda = \sqrt{2\pi}$, we can normalise it and define $f : x \longmapsto \frac{1}{\sqrt{2\pi}} q(x)$ which is again a nonnegative measurable function on $E = \mathbb{R}$. And it has the property $\int f d\lambda = 1$. Then, we can let :

$$\Phi(A) = \int_A f d\lambda = \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

to obtain a probability measure on $E = \mathbb{R}$. We now define a random variable $X : (\mathbb{R}, \mathcal{B}(\mathbb{R}), P) \longmapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, where P is any probability measure, by :

$$P(X \in A) = \Phi(A)$$

for all $A \in \mathcal{B}(\mathbb{R})$. X is said to be a **standard normal random variable** and we note $X \sim \mathcal{N}(0, 1)$.

Theorem 23. *If ν has density f with respect to μ, then :*

$$\int g d\nu = \int g f d\mu$$

Proof. See [5], p214. □

Generally, if we consider a random variable $X : (\Omega, \mathcal{F}, P) \longmapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ that has distribution μ, we say that it has **density** f with respect to λ if f is a non negative λ-measurable function and :

$$P(X \in A) = \mu(A) = \int_A d\mu = \int_A f(x)dx.$$

Taking $A = \mathbb{R}$, we necessarily have that f must integrate to 1 or $\int f(x)dx = 1$.

A.2 Convergence

Let $X, X_1, X_2, ...$ be random variables on $(\Omega, \mathcal{F}, P)$ taking their values in $\mathbb{R}^d$.

We say that the sequence (X_n) converges **almost surely** (or a.s.) to X, and we note $X_n \xrightarrow{a.s.} X$, if :

$$P(\omega \in \Omega \mid X(\omega) = \lim X_n(\omega)) = 1$$

and, for $p \in [1, \infty)$, we say that the sequence (X_n) converges in L^p to X, and we note $X_n \xrightarrow{L^p} X$, if :

$$\lim E(|X_n - X|^p) = 0.$$

These two types of convergence are the strongest. They imply the following type of convergence.

We say that the sequence (X_n) converges **in probability** to X, and we note $X_n \xrightarrow{P} X$, if :

$$\forall \epsilon > 0 \qquad \lim P(|X_n - X| > \epsilon) = 0.$$

Finally, we say that the sequence (X_n) converges **in distribtion** to X, and we note $X_n \xrightarrow{D} X$, if :

$$\lim P(X_n \leq x) = P(X \leq x) \qquad \forall x \in \mathbb{R} \text{ so that } F : y \longmapsto P(X \leq y) \text{ is continuous at } x.$$

A.3 Independence

We consider a probability space $(\Omega, \mathcal{F}, P)$. Two events A and B in $\mathcal{F}$ are **independent** if :

$$P(A \cap B) = P(A)P(B)$$

More generally, the events $A_1, A_2..., A_n$ are independent, if for all subsequence (j_k) of $\{1, 2, ..., n\}$, we have :

$$P(A_{j_1} \cap A_{j_2} \cap ... \cap A_{j_p}) = P(A_{j_1})P(A_{j_2})...P(A_{j_p}).$$

Note that this definition is not a simple extension of the case $n = 2$ and that it is *not* sufficient, for independence, to have the independence of any pair of events choosen in the collection $A_1, A_2, ..., A_n$. Indeed, consider the experiment consisting in tossing a (fair) coin two times. A, B, C being the events "getting a tail in the first trial", "getting a tail in the second trials" and "having the same result in both trial" respectively. These events are not independent although every pair of events choosen among them is.

Lemma 24. Let's consider the n events $A_1, A_2, ..., A_n$. We have that they are independent if and only if :

$$P(B_1 \cap B_1 \cap ... \cap B_n) = P(B_1)P(B_2)...P(B_n)$$

when $B_i \in \sigma(A_i) = \{\emptyset, A_i, A_i^c, \Omega\}$.

Proof. See [3], p110. □

The collections of $\mathcal{F}$-measurable sets $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n$, are independent if whenever $A_i \in \mathcal{A}_i$ and $I \subset \{1, ..., n\}$ we have $P(\bigcap_i A_i) = \prod_i P(A_i)$. An important special case appears when $\mathcal{A}_i$ are σ-algebras.

Let $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_n$ be subσ-algebras of $\mathcal{F}$. We say that $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_n$ are independent if and only if :

$$\forall A_1 \in \mathcal{F}_1, \forall A_2 \in \mathcal{F}_2, ..., \forall A_n \in \mathcal{F}_n \text{ we have } P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2)...P(A_n).$$

It follows naturally that n random variables $X_1, X_2, ..., X_n$ taking their values in $(E_1, \mathcal{E}_1), (E_2, \mathcal{E}_2), ..., (E_n, \mathcal{E}_n)$ respectively are said to be independent if the σ-algebras $\sigma(X_1), \sigma(X_2), ..., \sigma(X_n)$ are independent. That is :

$$\begin{aligned} \forall F_1 \in \mathcal{E}_1, \forall F_2 \in \mathcal{E}_2, ..., \forall F_n \in \mathcal{E}_n \text{ we have} \quad & P(\{X_1 \in F_1\} \cap \{X_2 \in F_2\} \cap ... \cap \{X_n \in F_n\}) \\ & = P(X_1 \in F_1)P(X_2 \in F_2)...P(X_n \in F_n). \end{aligned}$$

Theorem 25. *Suppose $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n$ are independent and each $\mathcal{A}_i$ is a π-system. Then $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2), ..., \sigma(\mathcal{A}_1)$ are independent.*

Proof. See [2], chapter 2. □

Theorem 26. *In order for $X_1, X_2, ..., X_n$ to be independent, it is sufficient that for all $x_1, x_2, ..., x_n \in (-\infty, \infty]$*

$$P(\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\} \cap ... \cap \{X_n \leq x_n\}) = \prod_{i=1}^{n} P(X_i \leq x_i)$$

Proof. See [2], p 44. □

Theorem 27. *Suppose $X_1, X_2, ..., X_n$ are independent random variables and that X_i has distribution μ_i. Then $(X_1, X_2, ..., X_n)$ has distribution $\mu_1 \times \mu_2 \times .... \times \mu_n$ that is :*

$$\begin{aligned} P((X_1, X_2, ..., X_n) \in A_1 \times A_2 \times ... \times A_n) \quad & = \mu_1(A_1)\mu_2(A_2)...\mu_n(A_n) \\ & = \mu_1 \times \mu_2 \times .... \times \mu_n(A_1 \times A_2 \times ... \times A_n) \end{aligned}$$

where $A_i \in \mathcal{B}(\mathbb{R})$ for all $i = 1, 2, ..., n$.

Proof. See [2], p 46. □

Theorem 28. *Suppose X, Y are independent random variables that have distributions μ and ν respectively. If $h : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is a measurable function with $h \geq 0$ or $E(|h(X, Y)|) < \infty$ then :*

$$E(h(X, Y)) = \int \int h(x, y)\mu(dx)\nu(dy)$$

Proof. See [2], p 46. □

In particular, if $h(x, y) = f(x)g(y)$ where f, g are measurable functions with $f, g \geq 0$ or $E(|f(X)|) < \infty$ and $E(|g(Y)|) < \infty$ then :

$$E(f(X)g(Y)) = E(f(X)) \cdot E(g(Y))$$

Note that by induction, we can extend this result to any finite sequence of independent random variables.

A.4 Gaussian Spaces

A random variable $X : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})$ that has a density f is called a **reduced Gaussian** random variable if :

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$$

and we write $X \sim \mathcal{N}(0,1)$ since then $E(X) = 0$ and $\mathrm{Var}(X) = 1$. Generally, we note $Y \sim \mathcal{N}(\mu, \sigma^2)$ when $E(Y) = \mu$ and $\mathrm{Var}(Y) = \sigma^2$. In this case, we must have $Y = \sigma X + m$ and the density f_Y of Y is given by:

$$f_Y(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}).$$

For any random variable X, we define its moment **characteristic function** by :

$$\phi(t) = E(e^{itX}).$$

The presence of the complex number i ensures the existence of ϕ. In the case of a Gaussian variable with mean μ and variance σ^2, we have :

$$\phi_Y(t) = \exp(it\mu - \frac{\sigma^2 t^2}{2}) \qquad\qquad Y \sim \mathcal{N}(\mu, \sigma^2).$$

The construction of Brownian motion will rely, partly, of the following theorem.

Theorem 29. *If (X_n) is a sequence of Gaussian random variables which converges in probability to a random variable X then X is a Gaussian random variable, the family $\{|X_n|^p\}$ is uniformly integrable and the convergence holds in L^p for every $p \geq 1$. Besides, if $X_n \sim \mathcal{N}(m_n, \sigma_n^2)$ then $X \sim \mathcal{N}(m, \sigma^2)$ where $m = \lim m_n$ and $\sigma^2 = \lim \sigma_n^2$ exist.*

Proof. See [4], p4. □

A **Gaussian space** is a closed linear subspace of a space $L^2(\Omega, \mathcal{F}, P)$ consisting *only* of centered ($\mu = 0$) Gaussian random variables. The existence of this space is given by the previous theorem.

Example 2. If we consider an euclidean space E of dimension n (similar to $\mathbb{R}^n$) and an inner product $\langle ., . \rangle$ on this space, then a random vector is a **Gaussian vector** if for all $u \in E$, we have that $\langle u, X \rangle$ is a Gaussian variable. Now, for any Gaussian vector $X = (X_1, X_2, ..., X_n)$ in $\mathbb{R}^n$, $\mathrm{span}\{X_1, X_2, ..., X_n\}$ is a Gaussian space.

A family $(X_t)_{t \in T}$ of real random variables is a (centered) **Gaussian process** if all the finite linear combinations of its elements are centered Gaussians.

Lemma 30. If $X = (X_t)_{t \in T}$ is a Gaussian process, then the closed vectorial subspace of L^2 generated by the random variables X_t, for $t \in T$ is a Gaussian space, called the **Gaussian space generated by the Gaussian process** X.

Proof. See [4], p9. □

Theorem 31. *Let $G_i, i \in I$ be a family of closed subspaces of a given Gaussian space; then, the σ-algebras $\sigma(G_i), i \in I$ are independent if and only if the spaces $G_i, i \in I$ are pairwise orthogonal.*

Proof. See [4], p9. □

Bibliography

[1] D. REVUZ., M. YOR *Continuous Martingales and Brownian Motion*, Springer, 3rd edition, 1999.

[2] R. DURRETT. *Probability: theory and examples*, Cambridge, 4th edition, 2010.

[3] J-F. LE GALL. *Intégration, Probabilités et Processus Aléatoires*, ENS Paris, 2006.

[4] J-F. LE GALL. *Mouvement Brownien et Calcul Stochastique*, Université Paris-Sud, 2008.

[5] P.BILLINGSLEY *Probability and Measure*, Wiley, 3rd edition, 1995.

[6] C. SHALIZI *Stochastic Processes (Advanced Probability II), 36-754*, Carnegie Mellon University, spring 2007.